Effective Health Care Program

Design and Implementation of N-of-1 Trials: A User's Guide

N of 1

Department of Health & Human Services • USA

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

The Agency for Healthcare Research and Quality's (AHRQ) Effective Health Care Program conducts and supports research focused on the outcomes, effectiveness, comparative clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services. More information on the Effective Health Care Program and electronic copies of this report can be found at www.effectivehealthcare.ahrq.gov.

This report was produced under contract to AHRQ by the Brigham and Women's Hospital DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) Methods Center under Contract No. 290-2005-0016-I. The AHRQ Task Order Officer for this project was Parivash Nourjah, Ph.D. The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ or the U.S. Department of Health and Human Services. Therefore, no statement in this report should be construed as an official position of AHRQ or the U.S. Department of Health and Human Services.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

The following investigators acknowledge financial support:

Dr. Naihua Duan was supported in part by NIH grants 1 R01 NR013938 and 7 P30 MH090322, Dr. Heather Kaplan is part of the investigator team that designed and developed the MyIBD platform, and this work was supported in part by Cincinnati Children's Hospital Medical Center. Dr. Richard Kravitz was supported in part by National Institute of Nursing Research Grant R01 NR01393801. Dr. Ida Sim is a co-investigator on an NINR R01 with Dr. Richard Kravitz (R01-NR13938) titled *N-of-1 Trials Using mHealth in Chronic Pain.* Dr. Sunita Vohra receives salary support as the recipient of an Alberta Innovates-Health Solutions Health Scholar Award.

Dr. Ian Eslick reports personal fees from Cincinnati Children's Hospital and Medical Center in support of the MyIBD system development. He owns an interest in the MyIBD technology through his company Vital Reactor LLC, which has plans to develop commercial offerings based on MyIBD. None of the other investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested Citation:

Kravitz RL, Duan N, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014. www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm.

Suggested citations for individual chapters are provided after the list of authors.

Design and Implementation of N-of-1 Trials: A User's Guide

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Prepared by:
Brigham and Women's Hospital DEcIDE Methods Center
Boston, MA

Contract No. 290-2005-0016-I

Editors:
Richard L. Kravitz, M.D., M.S.P.H.
Naihua Duan, Ph.D.

ISBN: 978-1-58763-430-7

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

AHRQ Publication No. 13(14)-EHC122-EF
February 2014

Acknowledgments

The editors and the DEcIDE Methods Center N-of-1 Guidance Panel would like to acknowledge Sebastian Schneeweiss, John D. Seeger, and Elizabeth Robinson Garry of the Brigham and Women's Hospital DEcIDE Methods Center, and Parivash Nourjah and Scott R. Smith of AHRQ for their guidance and assistance throughout this project. We consider ourselves fortunate to have worked with such smart, energetic, and committed colleagues.

Special thanks are due to Jiang Li (Chapter 1), Salima Punja (Chapter 2), and Elizabeth Staton (Chapter 3) for their contributions to the respective chapters, providing invaluable support to the Panel.

We also would like to thank Kristen Greenlee (UC Davis) for invaluable administrative support and the staff of AHRQ's Office of Communications and Knowledge Transfer (OCKT) for assistance with the editorial process. Their work occurred quietly in the background but was essential to the success of this project.

Preface

Design and Implementation of N-of-1 Trials: A User's Guide provides information on the design and implementation of n-of-1 trials (a.k.a. single-patient trials), a form of prospective research in which different treatments are evaluated in an individual patient over time. The apparent simplicity of this study design has caused it to be enthusiastically touted in some research fields and yet overlooked, underutilized, misunderstood, or erroneously implemented in other fields. With the advent of comparative effectiveness research and patient-centered outcomes research, there is a renewed interest in n-of-1 trials as an important research method for generating unique scientific evidence on patient health outcomes. A core aspect of this interest is that the n-of-1 approach may overcome some important limitations of other methodologies that involve larger samples of subjects. As a result, findings from n-of-1 trials may be especially useful in informing key health care decisions by patients and providers, particularly when combined with other scientific evidence. Likewise, the expansion of electronic health information technology into all areas of clinical care and the increasing recognition that new systems may also be deployed for research and quality improvement have further driven interest in conducting more n-of-1 trials as part of a learning health care system.[1]

AHRQ commissioned this User's Guide as an informational resource to researchers, health care providers, patients, and other stakeholders to improve general understanding of n-of-1 trials and strengthen the quality of evidence that is generated when an n-of-1 trial is conducted. The overarching aim of this User's Guide is to guide readers by identifying key decisions and tradeoffs in the design and implementation of n-of-1 trials, particularly when used for patient-centered outcomes research. Patient-centered outcomes research includes investigations of a wide range of research problems, particularly studying the outcomes, effectiveness, benefits, and harms of diagnostic tests, treatments, procedures, or health care services. This User's Guide identifies key elements to consider in applying the n-of-1 trial methodology to patient-centered outcomes research, describes some of the important complexities of the method, and provides readers with checklists to summarize the main points.

The production of this document was modeled on similar AHRQ initiatives to publish methods guides on topics such as systematic reviews,[2] medical tests,[3] patient registries,[4] and protocol development.[5] For this User's Guide, experts in the field of n-of-1 studies were identified and invited to participate in the development of the document as authors. Authors subsequently worked together to outline and write the document, which was subject to multiple internal and external independent reviews. All of the authors had the opportunity to discuss, review, and comment on the recommendations that are provided in this document, and these authors take responsibility for its scientific content.

Many individuals contributed to the production of this User's Guide and are acknowledged for their contributions to the project. Foremost among these are the authors and editors of the Guide. Each author has substantial expertise and experience with conducting or using n-of-1 trials and worked in an extremely thoughtful, collegial, and highly efficient manner to produce the Guide as a scholarly endeavor intended to benefit others in the researcher, patient, and health care practice communities.

This User's Guide would not have been possible without the leadership of Drs. Richard L. Kravitz and Naihua Duan, who served as editors of the Guide and authored chapters in it. Their scholarship in the field of n-of-1 trials, respected leadership, and hard work in all aspects of producing this document created an intellectually stimulating environment that inspired everyone who participated in the project. Likewise, Dr. Sebastian Schneeweiss, Dr. John D. Seeger, and Ms. Elizabeth Robinson Garry of Harvard Medical School and Brigham and Women's Hospital and Dr. Parivash Nourjah of AHRQ provided valuable scientific insights, technical assistance, and organizational support to ensure the successful publication of this User's Guide.

It is the hope of this entire team that researchers, clinicians, patients, and other stakeholders will find this User's Guide to be a valuable new resource in conducting patient-centered outcomes research. In particular, it is anticipated that:

- Investigators will find this document informative for planning and carrying out n-of-1 studies
- Clinicians will find this document useful for identifying patients who may benefit from participation in n-of-1 studies and informing them of the pros and cons of participation
- Patients who are active or prospective participants in n-of-1 studies will discover herein key concepts relevant to their participation and ultimate clinical decision; the document might be especially relevant to the emerging segment of patients who are interested in taking an active role in understanding their health outcomes and pathways to improvement
- Institutional Review Board (IRB) members and grant review boards will find this document useful to inform them on the goals and methods of n-of-1 trials, particularly the ethical aspects
- Health system administrators will find this document useful to inform them how to assess the value proposition for n-of-1 studies
- Statisticians will find useful information on how to provide statistical support to n-of-1 programs and
- Information technology (IT) directors and staff will find this document contains critical information on how to select hardware and software needed to support n-of-1 studies.

Undoubtedly, new approaches to n-of-1 trials will develop and the standards of practice will change or evolve over time, which will necessitate periodic updates of this User's Guide. Nonetheless, this document brings together the current knowledge of experts to lay the groundwork for designing and implementing high-quality n-of-1 trials for patient-centered outcomes research. As with other documents by AHRQ, this User's Guide is not intended to be prescriptive and is one of many resources that investigators and other stakeholders should consult when designing or appraising the results from an n-of-1 study. As new research methods, standards, and statistical tools develop, this User's Guide will need to be updated periodically in order to be useful to users. As a result, comments from investigators, stakeholders, and other users are welcome so they can be considered for incorporation into future versions of this User's Guide.

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Scott R. Smith, Ph.D.
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

1. Institute of Medicine. Best Care at Lower Cost: The Path to Continuously Learning Healthcare in America. Washington, DC: National Academy Press; 2012
2. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(12)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. April 2012. Chapters are available at www.effectivehealthcare.ahrq.gov.
3. Chang SM, Matchar DB, Smetana GW, et al., editors. Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012.
4. Gliklich RE, Dreyer NA, eds. Registries for Evaluating Patient Outcomes: A User's Guide. 2nd ed. (Prepared by Outcome DEcIDE Center [Outcome Sciences, Inc. d/b/a Outcome] under Contract No. HHSA29020050035I TO3.) AHRQ Publication No.10-EHC049. Rockville, MD: Agency for Healthcare Research and Quality. September 2010.
5. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. AHRQ Publication No. 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality; January 2013. www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm.

Contents

Chapter 1. Introduction to N-of-1 Trials: Indications and Barriers

Richard L. Kravitz, M.D., M.S.P.H.
University of California Davis

Naihua Duan, Ph.D.
Columbia University

Sunita Vohra, M.D.
University of Alberta

Jiang Li, M.D.
Lanzhou University

The DEcIDE Methods Center N-of-1 Guidance Panel*

Introduction

The goal of evidence-based medicine (EBM) is to integrate research evidence, clinical judgment, and patient preferences in a way that maximizes benefits and minimizes harms to the individual patient. The foundational, gold-standard research design in EBM is the randomized, parallel group clinical trial. However, the majority of patients may be ineligible for or unable to access such trials.[1] In addition, these clinical experiments generate average treatment effects, which may not apply to the individual patient; some patients may derive greater benefit than average from a particular treatment, others less. Patients want to know: which treatment is likely to work better for me? To generate individual treatment effects (ITEs), clinical investigators have taken several tacks, including subgroup analysis, matched pairs designs, and n-of-1 trials. Of these, n-of-1 trials provide the most direct route to estimating the effect of a treatment on the individual.

In this chapter, we introduce n-of-1 trials by providing definitions and a rationale, delineating indications for use, describing key design elements, and addressing major opportunities and challenges.

Defining N-of-1 Trials

N-of-1 trials in clinical medicine are multiple crossover trials, usually randomized and often blinded, conducted in a single patient. As such, n-of-1 trials are part of a family of Single Case Designs that have been widely used in psychology, education, and social work. In the schema of Perdices et al., the Single Case Designs family includes case descriptions, nonrandomized designs, and randomized designs.[2] N-of-1 trials are a specific form of randomized or balanced designs characterized by periodic switching from active treatment to placebo or between active treatments ("withdrawal-reversal" designs). N-of-1 trials were introduced to clinicians by Hogben and Sim as early as 1953,[3] but it took 30 years for the movement to find an effective evangelist in the person of Gordon Guyatt at McMaster University.[4-6] Many of the pioneers of the movement established active n-of-1 trial units in academic centers, only to abandon them once funding was exhausted.[7] However, several units are still thriving, and over the past three decades more than 2,000 patients have participated in published n-of-1 trials; fewer than 10 percent of them chose treatments inconsistent with the results.

In contrast to parallel group trials, n-of-1 trials use crossover between treatments to address the problem of patient-by-treatment interaction.

*Please see author list in the back of this User's Guide for a full listing of panel members and affiliations.

This situation arises when characteristics of the individual affect whether treatment A or treatment B (which could be an active treatment, a placebo, or no treatment) delivers superior results. Also, by prescribing multiple episodes of treatment, n-of-1 trials increase precision of measurement and control for treatment-by-time interaction, that is, the possibility that the relative effects of two treatments vary over time.

Rationale for N-of-1 Trials in the Era of Patient-Centered Care

The success of an n-of-1 trial largely depends on the collaboration and commitment of both clinician and patient. Clinicians must explain the process to their patients, collaborate with them in developing outcome measures most appropriate to the individual, monitor patients at regular intervals throughout the trial, evaluate and explain what the results of the trial mean, and work with patients to determine the course of treatment based on trial findings. Patients participating in n-of-1 trials must be involved in selecting therapies for evaluation, recording processes and outcomes (including nonadherence to treatment protocols), and sharing in treatment decisionmaking. As the centerpiece of patient-centered care, patient engagement has been shown to improve health outcomes among patients with chronic illness.[8] Nikles et al. reported that patients who had completed an n-of-1 trial had a greater understanding and awareness of their condition and felt a greater sense of control when it came to decisions about their health.[9]

It is the intent of this User's Guide to encourage and facilitate broader use of n-of-1 trials as a patient-centered clinical decision support tool. With appropriate infrastructure support, n-of-1 trials can be used by individual practicing clinicians in their daily care of individual patients. While the naturalistic application of n-of-1 trials may involve a single patient-clinician pair, over time there may emerge a multitude of such pairs. As discussed in the section "Statistical Analysis and Feedback for Decisionmaking," n-of-1 trials can be combined to provide more informative treatment decisions for individual patients by using information from other similar n-of-1 trials. This mimics the way that clinicians learn from their prior clinical experience and from their colleagues' clinical experience.

At the same time, research studies may use n-of-1 trials to examine decision support, quality improvement, and implementation of improved clinical and organizational procedures. As a research study design, n-of-1 trials are uniquely capable of informing clinical decisions for individual patients. Therefore, the research goal (to produce generalizable knowledge that can be applied to future patients) is compatible with the clinical goal of serving the needs of the individual patients participating in these trials. (The same is often not true for other research study designs, such as the usual parallel group randomized controlled trials [RCTs], in which patients contribute to the research but usually do not benefit directly in terms of their own clinical decisionmaking.) This special feature of n-of-1 trials may facilitate the recruitment and retention of patients and clinicians in research studies.

Beyond their potential for promoting patient-centered care, n-of-1 trials may have additional pragmatic value. With escalating drug costs, health care systems are struggling to provide cost-effective therapies. N-of-1 trials offer an objective way of determining individual response to therapy: if two therapeutic options are shown to have equivalent effectiveness in a given individual, the less costly option could be chosen. This approach to comparative effectiveness could apply to different classes of medications, as well as formal assessment of the bioequivalence of generic and proprietary pharmaceuticals. Considering that n-of-1 trials are particularly suited to chronic conditions, the savings to the health care system could be substantial.

Indications, Contraindications, and Limitations

N-of-1 trials are indicated whenever there is substantial uncertainty regarding the comparative effectiveness of treatments being considered for an individual patient. Uncertainty can result from a general lack of evidence (as when no relevant parallel group RCTs have been conducted), when the existing evidence is in conflict, or when the evidence is of questionable relevance to the patient at hand.[10] Uncertainty may also result from heterogeneity of treatment effects (HTE) across patients that cannot be easily predicted from available prognostic factors. HTE is the variance of ITEs across patients, where the ITE is the difference in effects (net benefits) between treatment A and treatment B for an individual patient.[11] Though the extent of HTE for common conditions and treatments is not well characterized, some analyses suggest it is substantial.[12-16]

N-of-1 trials are applicable to chronic, stable, or slowly progressive conditions that are either symptomatic or for which a valid biomarker has been identified. Acute conditions offer no opportunity for multiple crossovers. Rapidly progressive conditions (or those prone to sudden, catastrophic outcomes such as stroke or death) are not amenable to the deliberate experimentation of n-of-1 trials. Asymptomatic conditions make outcomes assessment difficult, unless a valid biomarker exists.[17] Examples of such biomarkers might include blood pressure or LDL cholesterol in heart disease, sedimentation rate in some chronic autoimmune diseases, or intraocular pressure in glaucoma. Some patient groups (e.g., patients with rare diseases) may be particularly motivated to participate in n-of-1 trials owing to the paucity of other evidence needed to substantiate treatment effect.

For practical reasons, treatments to be assessed in n-of-1 trials should have relatively rapid onset and washout (i.e., few lasting carryover effects). Treatments with a very slow onset of action (e.g., methotrexate in rheumatoid arthritis) could outlast the patience of the average patient and clinician.[18] On the other hand, treatments with prolonged carryover effects would require a substantial washout period to distinguish between the effects of the current treatment and the previous treatment.[5] In addition, regimens requiring complex dose titration (e.g., loop diuretics in patients with comorbid congestive heart failure and chronic kidney disease) are not well suited for n-of-1 trials.

Major Design Elements of N-of-1 Trials

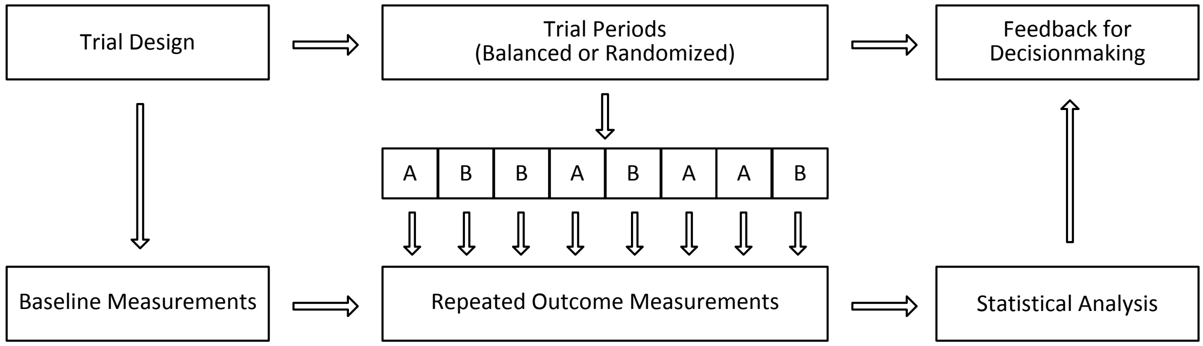
The major design elements of n-of-1 trials are balanced sequence assignment, blinding, and systematic outcomes measurement. Before introducing these elements, we offer a description of standard clinical practice.

Standard Clinical Practice

In ordinary practice, the clinician prescribes treatment and asks that the patient return for followup. At the followup encounter, the clinician asks the patient if he or she is improving. If the patient responds positively, the treatment is continued. If not, the clinician and patient discuss alternative strategies such as a dose increase, switching to a different treatment, or augmenting with a second treatment. This process continues until both agree that a satisfactory outcome has been achieved, until intolerable side effects occur, or until no further progress seems possible. Although treatments are administered in sequence, there is no systematic repetition of prior treatments (replication), and the treatment assignment sequence is based on physician and patient discretion (not randomized or balanced). Neither clinician nor patient is blinded. Typically, there is no systematic assessment of outcomes. As a result, it is easy for both patient and clinician to be misled about the true effects of a particular therapy.

Take for example Mr. J, who presents to Dr. Alveolus with a nagging dry cough of 2 months duration that is worse at night. After ruling out drug effects and infection, Dr. Alveolus posits perennial (vasomotor) rhinitis with postnasal drip as the cause of Mr. J's cough and prescribes diphenhydramine 25 mg each night. The patient returns in a week and notes that he's a little better, but the "cough is still there." Dr. Alveolus increases the diphenhydramine dose to 50 mg, but the patient retreats to the lower dose after 3 days because of intolerable morning drowsiness with the higher dose. He returns complaining of the same symptoms 2 weeks later; the doctor prescribes cetirizine 10 mg (a nonsedating antihistamine). Mr. J fills the prescription but doesn't return for followup until 6 months later because he feels better. "How did the second pill I prescribed work out for you," Dr. Alveolus asks. "I think it helped," Mr. J replies, "but after a while the cough seemed to get better so I stopped taking it. Now it's worse again, and I need a refill."

Figure 1–1. Scheme for a prototypical n-of-1 trial (modified from Zucker et al.[20])



While this typical clinical scenario involves some effort to learn from experience, the approach is rather haphazard and can be improved upon.

N-of-1 Trial Procedures in Contrast to Standard Clinical Practice

What if Mr. J and Dr. Alveolus were to acknowledge their uncertainty and elect to embark on an n-of-1 trial of diphenhydramine versus cetirizine for treatment of chronic cough presumed due to perennial rhinitis? They might agree:

- To administer diphenhydramine and cetirizine in a balanced sequence of 7-day treatment intervals[a] for a total of eight treatment periods (four periods on diphenhydramine, four periods on cetirizine, 56 days total), with no washout time between treatment periods;
- To ask the compounding pharmacist to place the medications in identical capsules; and
- To assess benefits using the average of Mr. J's rating of overall cough severity (1–5 scale) and Mrs. J's rating of nighttime cough severity (1–5 scale) and harms using a daytime sleepiness scale. At the end of the trial, tradeoffs between benefits (decreased cough) and harms (increased drowsiness) can be examined either implicitly (through mutual deliberation between clinician and patient) or explicitly (using shared decisionmaking tools that assign specific weights to particular benefits and harms).[19]

Their design (schematized in Figure 1–1, modified from Zucker et al.[20]) incorporates balanced sequence assignment, repetition, blinding, and systematic outcomes assessment, which we now discuss in greater detail.

Balanced Sequence Assignment

In parallel group RCTs, randomization serves to maximize the likelihood of equivalence between treatment groups (in terms of both known and unknown prognostic factors). In n-of-1 trials, the aim is to achieve balance in the assignment of treatments over time so that treatment effect estimates are unbiased by time-dependent confounders. Randomization of treatment periods is one way of achieving such balance, but there are others. For example, the treatment sequence AAAABBBB offers no protection against a confounder whose effect on the outcome is linear with time (e.g., a secular trend). The paired design ABABABAB and the singly counterbalanced design ABBAABBA offer better protection against temporally linear confounders but are still vulnerable to nonlinear confounding. The doubly counterbalanced design ABBABAAB defends against both linear secular trends and nonlinear trends.

Balanced assignment (which may be achieved using randomization) helps control for time-varying clinical and environmental factors that could affect the patient's outcome.[21,22] Some, but not all, of these factors may be known to the patient and clinician in advance. For example, Mr. J might have decided to take diphenhydramine on weekends and cetirizine on weekdays. He might

[a]A 7-day interval is chosen for convenience and because shorter intervals might introduce confounding by day-of-week effects (e.g., the difference between weekends and weekdays).

then be less prone to notice daytime sleepiness from diphenhydramine because he tends to sleep in on weekends. This would bias his assessment. Randomization (along with blinding) makes it more difficult to guess which treatment has been assigned.

Repetition

For a patient interested in selecting the treatment likely to work best for him or her in the long term, the simplest n-of-1 trial design is exposure to one treatment followed by the other (AB or BA). This simple design allows for direct comparison of treatments A and B and protects against several forms of systematic error (e.g., history, testing, regression to the mean).[23] However, one-time exposure to AB or BA offers limited protection against other forms of systematic error (particularly maturation and time-by-treatment interactions) and virtually no protection against random error. To defend against random error (the possibility that outcomes are affected by unmeasured, extraneous factors such as diet, social interactions, physical activity, stress, and the tendency of symptoms to wax and wane over time), the treatment sequences need to be repeated (ABAB, ABBA, ABABAB, ABBAABBA, etc.). In this way, repetition is to n-of-1 trials what sample size is to parallel group RCTs.

Washout and Run-In Periods

The importance of a washout period separating active treatment periods in n-of-1 trials has been fiercely debated.[24] A washout period is theoretically important whenever lingering effects of the first treatment might influence outcome measurements obtained while on a subsequent treatment. Carryover effects resulting from insufficient washout will often tend to reduce observed differences between treatments for placebo-controlled trials. However, more complex interactions are possible. For example, if the benefits of a particular treatment wash out quickly but the risks of adverse treatment-related harm persist (think aspirin, which reduces pain over a matter of hours but increases risk of bleeding for up to 7 days), the likelihood of detecting net benefit will depend on the order in which the treatments are administered. Similar issues also apply to slow onset of the new treatment. A possible downside of a washout period is that the patient is forced to spend some time completely off treatment, which might be undesirable for patients who already receive some benefit from both treatments. For practical purposes, washout periods may not be necessary when treatment effects (e.g., therapeutic half-lives) are short relative to the length of the treatment periods. Since treatment half-lives are often not well characterized and vary among individuals, the safest course may be to choose treatment lengths long enough to accommodate patients with longer than average treatment half-lives and to take frequent (e.g., daily) outcome measurements. An alternative to the use of a "physical" washout is the use of "analytic washout," that is, to address the effects of carryover and slow onset analytically. Further discussion is offered in Chapter 4 (Statistics).

Some n-of-1 investigators have advocated for the use of run-in periods. In parallel group RCTs, a run-in period is a specified period of time after enrollment and prior to randomization that is allotted to further measure a participant's eligibility and commitment to a study.[25] In n-of-1 trials, a run-in period could also be used to differentiate "responders" from "nonresponders" in an open-label (unblinded) situation or to initiate dose-finding.

Blinding

In parallel group RCTs, blinding of patients, clinicians, and outcomes assessors ("triple blinding") is considered good research practice. These trials aim to generate generalizable knowledge about the effects of treatment in a population. In drug and device trials, the consensus is that it is critical to separate the biological activity of the treatment from nonspecific (placebo) effects. (For a broader view, see Benedetti et al.[26]) In n-of-1 trials, the primary aim is usually different. Patients and clinicians participating in n-of-1 trials are likely interested in the net benefits of treatment overall, including both specific and nonspecific effects. Therefore blinding may be less critical in this context. Nevertheless, expert opinion tends to favor blinding in n-of-1 trials whenever feasible.

However, just as in parallel group randomized trials, blinding is not always feasible. For example, in trials of behavioral interventions (e.g., bibliotherapy versus computer-based cognitive behavioral therapy for depression), patients will always know what treatment they are on. Furthermore, even for drug trials, few community practitioners have access to a compounding pharmacy that can safely and securely prepare medications to be compared in matching capsules.

Systematic Outcomes Assessment

Evidence is accumulating that careful, systematic monitoring of clinical progress supports better treatment planning and leads to better outcomes. For example, home blood pressure monitoring results in better blood pressure control,[27] and "treat-to-target" approaches based on PHQ-9 scores have worked well in depression.[28] In n-of-1 trials, systematic assessment of outcomes may well be the single most important design element. There are two issues to consider: (1) what data to collect and (2) how to collect them.

In designing an n-of-1 trial, participants (patients, clinicians, investigators) must first select outcome domains (specific symptoms, specific dimensions of health status, etc.) and then specific measures tapping those domains. In so doing, they must balance a number of competing interests. For most chronic conditions, there are numerous potentially relevant outcomes. These may be condition specific (e.g., pain intensity in chronic low back pain, diarrhea frequency in inflammatory bowel disease) or generic (e.g., health-related quality of life). Clinicians, patients, and service administrators may assign different priorities to different domains. For example, in chronic musculoskeletal pain, the patient may prioritize control of pain intensity or fatigue, the clinician may prioritize daily functioning, and Drug Enforcement Agency officials may prioritize minimizing opportunities for misuse of opiates. The primary purpose of most n-of-1 trials is to assist with individual treatment decisions. Therefore patient preferences are paramount. However, as prescribers of treatment, clinicians are essential partners, and their buy-in is essential.

Once outcome domains have been identified, participants need to pick specific measures. Though measures known to possess high reliability and validity are preferable, sometimes an appropriate pre-existing measure cannot be found. In this case, n-of-1 participants must choose between measures that are well validated but imprecisely targeted to the patient's goals or new measures that are incompletely validated but a good fit with patient priorities. An interesting compromise is a validated questionnaire (e.g., Measure Your Medical Outcome Profile, or MYMOP) that uses standardized wording and response options applied to the symptoms and concerns of greatest interest to the patient.[29]

N-of-1 trials can make use of the entire spectrum of data-collection modalities. Traditional approaches include surveys, diaries, medical records, and administrative data. Recent developments in information technology have opened the door to several new approaches, including ecological momentary assessment (EMA) and remote positional and physiologic monitoring. Mobile-device EMA cues the patient to input data at more frequent intervals (e.g., hourly, daily, or weekly) than is typical using traditional survey modalities. Compliance with such devices is higher than with paper diaries.[30] When equipped with GPS or movement detection technology (actigraphy), mobile devices can also track patient movements and activities. Ancillary monitoring devices can be connected to mobile devices to monitor heart rate, blood pressure, blood glucose, galvanic skin response, electroencephalographic activity, degree of social networking, vocal stress, etc. Data on the reliability and validity of these measures are currently scant but are accumulating rapidly.[31]

Statistical Analysis and Feedback for Decisionmaking

Once data are collected, they need to be analyzed and presented to the relevant decisionmakers in a format that is actionable. In the systematic review by Gabler et al.,[32] approximately half of the trials reported using a t-test or other simple statistical criterion (44%), while 52 percent reported using a visual/graphical comparison alone. Of the 60 trials (56%) reporting on more than one individual, 26 trials (43%) reported on a pooled analysis. Of these, 23 percent used Bayesian methodology, while the rest used frequentist approaches to combining the data. Guidance on statistical analytic approaches for n-of-1 trials is provided in Chapter 4 (Statistics).

While n-of-1 trials can promote other goals (e.g., increased patient engagement),[9] the primary objective is generally to promote better health care decisionmaking for participating patients. The degree to which decisionmaking can be improved will depend on the quality of the data (as influenced by trial design and measurement instruments) and the clarity with which results are communicated to the end-users, especially the patient participating in the trial. There are three fundamental issues n-of-1 trialists should consider. First, should outcomes data be presented item by item (or scale by scale) or as a composite measure? A patient with asthma may be interested in her ability to climb stairs, sleep through the night, and avoid the emergency room. These outcomes could be presented as three separate statistics, graphs, or figures, or they could be combined into a single composite measure that averages the individual components, possibly weighted to reflect the relative importance of the respective components. The advantage of single measures is that they retain clinical granularity and, in and of themselves, are readily interpretable. The disadvantage is that they can be confusing, especially if multiple outcomes are affected differently by the treatments under study. The advantage of composite measures is that they make individual-level decisionmaking more straightforward. If, for a given patient, the Asthma Improvement Index moves in a more positive direction on treatment A than B, the drug of choice is treatment A. The composite measure directly (if arbitrarily) addresses the tradeoff among the components such as benefits and harm, especially when the components respond to the treatments differently. On the other hand, composite outcomes are harder to interpret and may be driven by the most sensitive component (which is not necessarily the most important).

The second issue is how to present the data: as graphics, statistics, or both. Simple graphical analysis can transmit results clearly, but not all formats are equally understandable, particularly to low-numeracy populations.[33] In addition, graphical analysis can magnify small differences that a proper statistical analysis would show are likely due to chance. A combined approach may work best, employing statistics to test for stochastic significance (or, using a Bayesian framework, to estimate post-test probabilities) and graphics to lend clarity to the findings.

The third issue is whether to rely solely on the results of the current n-of-1 trial for decisionmaking or to "borrow from strength" by combining current data with the results of previous n-of-1 trials completed by similar patients. The choice will usually be driven by the availability of relevant data and by the ratio of within-patient versus between-patient variance (see Chapter 4 for details). If a similar series of trials has never been conducted, and if few patients have been enrolled in the current series, then decisionmaking rests by default on the results of the current n-of-1 trial alone. If, on the other hand, large numbers of patients have completed similar n-of-1 trials, and if within-patient variance is larger than between-patient variance, then "borrowing from strength" will enhance the precision of the result. Similar considerations influence the decision whether and how to combine current n-of-1 results with results extracted from the existing population evidence base (RCTs and observational studies). Further discussion is presented in Chapter 4 (Statistics).

Opportunities and Challenges

In addition to their potential for enhancing therapeutic precision, n-of-1 trials may offer three broader benefits. First, they may help patients and clinicians recognize ineffective therapies, thus reducing polypharmacy, minimizing adverse effects, and conserving health care resources. For example, if the marginal benefits of a new therapy were shown to be small, patient and clinician might elect to use the nearly equivalent but less costly therapeutic alternative. Second, they may help engage patients in their own care.[9] A robust literature supports the premise that increased patient involvement in care is associated with better outcomes.[8,34] By helping patients attend to their own outcomes and think critically about treatments, n-of-1 trials can awaken patients' "inner scientist" and give them a greater stake in the process of clinical care. Third, n-of-1 trials can blur the boundaries between clinical practice and clinical research, making research more like practice and practice more like research. Making research more like practice is desirable to increase the relevance and generalizability of clinical research findings. On the other hand, making practice more like research will create opportunities for developing the clinical evidence base by enhancing systematic data collection on the comparative effectiveness of treatments by real health care professionals treating real patients. As n-of-1 trials become better integrated into practice, the downstream benefits may include:

- Patients become more acquainted with the scientific method and in particular the value of rigorous clinical experiments.
- Clinicians become more connected to the process of generating clinical evidence, more engaged in clinical research, and potentially more interested in participating in clinical trials.
- Practices start collecting data on the relationship between treatments and outcomes and making such data available for use in routine patient care. If leveraged to full advantage, these data could become the linchpin of a "learning health care system" as envisioned by the Institute of Medicine.[35]

For such benefits to be realized, however, a number of challenges must be overcome. Most importantly, a business case must emerge that leaves patients, clinicians, and health care organizations convinced that increased therapeutic precision afforded by n-of-1 trials is worth the trouble. In addition, institutional ethics boards need to accept n-of-1 trials as an extension of clinical care; statistical procedures for the design and analysis of n-of-1 trials need to be automated into user-friendly tools accessible to clinicians and patients; health informatics systems must be created to support the seamless integration of n-of-1 trials into clinicians' practices and patients' lives; and all those concerned with improving the quality of therapeutic decisionmaking need adequate training and support. These topics are taken up in the remainder of the User's Guide.

Outline for the Rest of the User's Guide

In the rest of this User's Guide, authors will expand on the themes introduced here. Chapter 2 addresses human subjects issues germane to n-of-1 trials, in particular how n-of-1 trials are situated on the continuum between clinical care and research and hybrids in between. This chapter also provides guidance for IRB committees considering applications to conduct n-of-1 trials. Chapter 3 takes on the very practical issue of how much n-of-1 trials cost, how much value they offer, and what factors organizations should consider before constructing or endorsing an n-of-1 trial service. Chapter 4 provides an overview of statistical design and analysis considerations, while Chapter 5 outlines key components of information technology infrastructure needed to deploy n-of-1 trials efficiently. Finally, Chapter 6 takes up training and engagement of clinicians and patients preparing to participate in n-of-1 trials.

Checklist

Guidance	Key Considerations	Check
Determine whether n-of-1 methodology is applicable to the clinical question of interest	• Indications include: (a) substantial clinical uncertainty; (b) chronic or frequently recurring symptomatic condition; (c) treatment with rapid onset and minimal carryover. • Contraindications include: (a) rapidly progressive condition; (b) treatment with slow onset or prolonged carryover; (c) patient or clinician insufficiently interested in reducing therapeutic uncertainty to justify effort.	❒
Select trial duration, treatment period length, and sequencing scheme	• Longer trial duration delivers greater precision, but completion can be difficult or tedious, with the potential for extended exposure to inferior treatment during trial. • Treatment period length should be adjusted to fit the therapeutic half-life (of drug treatments) or treatment onset and duration (of nondrug treatments). • Simple randomization (e.g., AABABBBA) optimizes blinding (more difficult to guess treatment), while balanced sequencing (e.g., ABBABAAB) is a more reliable guarantor of validity.	❒
Invoke a suitable washout period, if indicated	• Washout is not necessary if treatment duration of action is short relative to treatment period. • Washout is contraindicated if patient could be harmed by cessation of active treatment.	❒
Decide whether or not to invoke blinding	• Blinding is feasible for some drug treatments but infeasible for most nondrug treatments (behavioral, lifestyle). • Adequate blinding allows investigators to distinguish between specific and nonspecific treatment effects. • In some circumstances, this distinction may not matter to patient and clinician; in others, participants may be primarily interested in the combined treatment effect (specific + nonspecific).	❒
Select suitable outcomes domains and measures	• Patient preferences are preeminent, but clinicians' goals and external factors should be accounted for and may occasionally supervene. • Valid and reliable measures are preferred when available, but patient-centeredness should not be sacrificed to psychometric imperatives.	❒
Analyze and present data to support clinical decisionmaking by patients and clinicians	• There is a natural tension between identifying a single, primary outcome for decisionmaking and coming to a full understanding of the data. • A reasonable approach is to select one or two primary outcome measures but present or use a variety of statistical and graphical methods to fully explicate the data.	❒

References

1. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. JAMA. Sep 13 2000;284(10):1290-1296.
2. Perdices M, Tate RL. Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? Neuropsychological rehabilitation. Dec 2009;19(6):904-927.
3. Hogben L, Sim M. The self-controlled and self-recorded clinical trial for low-grade morbidity. British Journal of Preventive and Social Medicine. Oct 1953;7(4):163-179.
4. Guyatt G, Taylor DW, Chong J, et al. Determining optimal therapy–randomized trials in individual patients. N Engl J Med. 1986;314(14):889-892.
5. Guyatt G, Adachi J, Roberts R, et al. A clinician's guide for conducting randomized trials in individual patients. Canadian Medical Association Journal (CMAJ) 1988 Sep 15;139(6):497-503.
6. Guyatt GH, Heyting A, Jaeschke R, et al. N of 1 randomized trials for investigating new drugs. Control Clin Trials. 1990 Apr;11(2):88-100.
7. Kravitz RL, Niedzinski EJ, Hay MC, Subramanian SK, Weisner TS. What ever happened to N-of-1 trials? Insiders' perspectives and a look to the future. Milbank Q. 2008 Dec;86(4):533-555.
8. Greenfield S, Kaplan S, Ware JE, Jr. Expanding patient involvement in care. Effects on patient outcomes. Annals of Internal Medicine. Apr 1985;102(4):520-528.
9. Nikles CJ, Del Mar CB. Using n-of-1 trials as a clinical tool to improve prescribing. Br J Gen Pract. 2005 Mar;55(512):175-180.
10. Greenfield S, Duan N, Kaplan SH. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. Am J Med. 2007 Apr;120(4 Suppl 1):S3-9.
11. Kravitz RL, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Q. 2004;82(4):661-687.
12. Kent DM, Rothwell PM, Ioannidis JP, et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials. 2010;11:85.
13. Rothwell PM, Mehta Z, Howard SC, et al. Treating individuals 3: from subgroups to individuals: general principles and the example of carotid endarterectomy. Lancet. Jan 15-21 2005;365(9455):256-265.
14. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" Lancet. Jan 1-7 2005;365(9453):82-93.
15. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet. Jan 8-14 2005;365(9454):176-186.
16. Warden D, Rush AJ, Trivedi MH, et al. The STAR*D Project results: a comprehensive review of findings. Current Psychiatry Reports. Dec 2007;9(6):449-459.
17. De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. summary of a National Institutes of Health workshop. Controlled clinical trials. Oct 2001;22(5):485-502.
18. Kravitz RL, White RH. N-of-1 Trials of Expensive Biological Therapies: A ThirdWay? . Archives of Internal Medicine 2008;168:1030–1033.
19. O'Connor AM, Legare F, Stacey D. Risk communication in practice: the contribution of decision aids. BMJ. Sep 27 2003;327(7417):736-740.
20. Zucker DR, Ruthazer R, Schmid CH, et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. Journal of Rheumatology. 2006;33(10):2069-2077.
21. Senn S. Cross-over trials in clinical research. 1993.
22. Senn S. Individual response to treatment: is it a valid assumption? BMJ. 2004 Oct 23;329(7472):966-968.
23. Campbell DT, Stanley J. Experimental and Quasi-Experimental Designs for Research. Boston: Houghton Mifflin Company; 1966.
24. Matthews JN. Multi-period crossover trials. Statistical methods in medical research. Dec 1994;3(4):383-405.
25. Ulmer M, Robinaugh D, Friedberg JP, et al. Usefulness of a run-in period to reduce drop-outs in a randomized controlled trial of a behavioral intervention. Contemporary clinical trials. Sep 2008;29(5):705-710.

26. Benedetti F, Mayberg HS, Wager TD, et al. Neurobiological mechanisms of the placebo effect. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. Nov 9 2005;25(45):10390-10402.

27. Agarwal R, Bills JE, Hecht TJ, et al. Role of home blood pressure monitoring in overcoming therapeutic inertia and improving hypertension control: a systematic review and meta-analysis. Hypertension. Jan 2011;57(1):29-38.

28. Lin EH, Von Korff M, Ciechanowski P, et al. Treatment adjustment and medication adherence for complex patients with diabetes, heart disease, and depression: a randomized controlled trial. Annals of Family Medicine. Jan-Feb 2012;10(1):6-14.

29. Paterson C. Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. BMJ. Apr 20 1996;312(7037):1016-1020.

30. Stone AA, Shiffman S, Schwartz JE, et al. Patient compliance with paper and electronic diaries. Controlled Clinical Trials. 04/ 2003;24(2):182-199.

31. Ramanathan N, Alquaddoomi F, Falaki H, et al. ohmage: An open mobile system for activity and experience sampling. Paper presented at: Pervasive Computing Technologies for Healthcare (Pervasive Health), 2012 6th International Conference on 21-24 May 2012, 2012.

32. Gabler NB, Duan N, Vohra S, et al. N-of-1 trials in the medical literature: a systematic review. Medical Care. Aug 2011;49(8):761-768.

33. Ancker JS, Senathirajah Y, Kukafka R, et al. Design Features of Graphs in Health Risk Communication: A Systematic Review. Journal of the American Medical Informatics Association. November 1, 2006 2006;13(6):608-618.

34. Bodenheimer T, Lorig K, Holman H, et al. Patient self-management of chronic disease in primary care. JAMA. Nov 20 2002;288(19):2469-2475.

35. Institute of Medicine. Best care at lower cost: the path to continuously learning healthcare in America. Washington, DC: National Academy Press; 2012.

Chapter 2. An Ethical Framework for N-of-1 Trials: Clinical Care, Quality Improvement, or Human Subjects Research?

Salima Punja, B.Sc., Ph.D. Candidate
Sunita Vohra, M.D., M.R.C.P.C., M.Sc.
University of Alberta

Ian Eslick, Ph.D.
MIT Media Laboratory

Naihua Duan, Ph.D.
Columbia University

The DEcIDE Methods Center N-of-1 Guidance Panel*

Introduction

N-of-1 trials are prospectively planned multiple crossover trials conducted in a single individual.[1] They can be used to evaluate a wide range of conditions including neurological, behavioral, rheumatologic, pulmonary, and gastrointestinal conditions. They are useful when the patient's symptoms are stable (or frequently occurring), and the treatment takes effect quickly, with few or no residual carryover effects. Further discussion concerning the features and indications for these trials can be found in Chapter 1 (Introduction) of this User's Guide.

Whether n-of-1 trials are a form of systematic learning with the aim of promoting evidence-based clinical care (and therefore a form of "quality improvement") or experiments that aim to produce generalizable knowledge (and therefore "research"), depends on the intent of the trial. In this chapter, we will consider issues that influence whether n-of-1 trials should be treated as clinical care or research. Settling this question is a critical prelude to addressing a number of major ethical concerns and providing guidance for discussions with institutional research ethics boards. We start by considering how n-of-1 trials compare to trials of therapy in routine clinical practice and to traditional randomized controlled trials (RCTs).

N-of-1 Trials Compared to "Trials of Therapy" in Usual Care

"Trials of therapy" or "therapeutic trials" have been utilized extensively in clinical practice to evaluate therapeutic effectiveness in individual patients for a wide variety of therapies, including medication (such as dose determination), device, behavioral and lifestyle therapies, etc. Such informal trials are part of usual care, are unblinded, have no control conditions, and involve no formal validated assessment of effectiveness. As a result, they are vulnerable to bias and uncertainty. Unlike trials of therapy, n-of-1 trials utilize multiple comparisons with a control condition (active or placebo) and a priori decisions about choice and timing of outcome assessment. As such, n-of-1 trials (compared to informal trials of therapy) reduce the risk of drawing invalid conclusions about the effectiveness of a therapy in an individual patient. More specifically, both informal trials of therapy and n-of-1 trials can be utilized in clinical care; however, n-of-1 trials can lead to better therapeutic decisions and outcomes.

*Please see author list in the back of this User's Guide for a full listing of panel members and affiliations.

N-of-1 Trials Compared to Traditional RCTs

RCTs, the gold standard of clinical research, protect against bias by utilizing blinding, randomization, control conditions, and a priori decisions about outcome measure assessment. While these elements protect internal validity, typical parallel group RCTs have been criticized for their limited external validity and generalizability.[2] For example, restrictive inclusion and exclusion criteria may limit RCT enrollment to less than 10 percent of individuals with the disease in question.[3] Unlike the usual parallel group RCTs, n-of-1 trials can be tailored to the condition and treatment in question, as well as the outcomes most relevant to the patient. As a result, it has been suggested that the n-of-1 trial design has the potential to provide the strongest evidence for individual treatment decisions and should therefore occupy the pinnacle of the evidence pyramid.[4] Furthermore, a series of n-of-1 trials testing the same intervention and conducted in similar patients with identical outcome measures may be pooled for meta-analysis, potentially generating estimates of treatment effect that are relevant for a population.[5] Thus, although n-of-1 trials can be utilized as individualized trials of therapy in clinical care settings, the same trial design can also be utilized as a research tool to extend the scope of the usual parallel group RCTs.

N-of-1 Trials: Clinical Care Versus Clinical Research

Differentiating clinical care from research employing experimental therapies can be difficult.6 Quite apart from research, clinical innovation may involve the use of novel therapies or existing therapies for new indications. The application of these therapies is determined by clinical judgment and overseen by all the usual channels for supervising clinical patient care.

Research with experimental therapy can also involve a single individual and is administered by the researcher, preferably in close collaboration with clinical expert(s), and is overseen by institutional research ethics boards. Careful consideration of the features that distinguish clinical innovation from research is needed, especially in the context of chronic disease management (see Table 2–1 for further consideration of the differences between clinical care, n-of-1 trials, and clinical research).

Discourse on the ethics of n-of-1 trials depends on the trial's intention: research versus learning for clinical care. For example, in research, the goal is to produce a generalizable result; any benefit gained by individual participants is secondary. In clinical care, however, the primary goal is to determine treatment effectiveness for the individual patient. The two activities are fundamentally different in their intent and therefore require different ethical considerations.

More specifically, the U.S. Department of Health and Human Services' Office for Human Research Protections (HHS/OHRP) defines research that is subject to human subjects regulations as follows:

> *Research* means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy, whether or not they are conducted or supported under a program which is considered research for other purposes. For example, some demonstration and service programs may include research activities.[7]

Using this definition, n-of-1 trials designed to evaluate therapeutic effectiveness in a single individual are not research (Figure 2–1).

The U.S. HHS/OHRP clarifies the distinction between human subjects research and quality improvement for clinical care as follows:

> Protecting human subjects during research activities is critical and has been at the forefront of HHS activities for decades. In addition, HHS is committed to taking every appropriate opportunity to measure and improve the quality of care for patients. These two important goals typically do not intersect, since most quality improvement efforts are not research subject to the HHS protection of human subjects regulations. However, in some cases quality improvement activities are designed to accomplish a research purpose as well as the purpose of improving the quality of

care, and in these cases the regulations for the protection of subjects in research (45 CFR part 46) may apply (HHS, 2009).

To determine whether these regulations apply to a particular quality improvement activity, the following questions should be addressed in order:

1. Does the activity involve *research* (45 CFR 46.102(d))?
2. Does the research activity involve human subjects (45 CFR 46.102(f))?
3. Does the human subjects research qualify for an exemption (45 CFR 46.101(b))?
4. Is the nonexempt human subjects research conducted or supported by HHS or otherwise covered by an applicable FWA approved by OHRP?

For those quality improvement activities that are subject to these regulations, the regulations provide great flexibility in how the regulated community can comply. Other laws or regulations may apply to quality improvement activities independent of whether the HHS regulations for the protection of human subjects in research apply (HHS, 2009: http://answers.hhs.gov/ohrp/questions/7281).

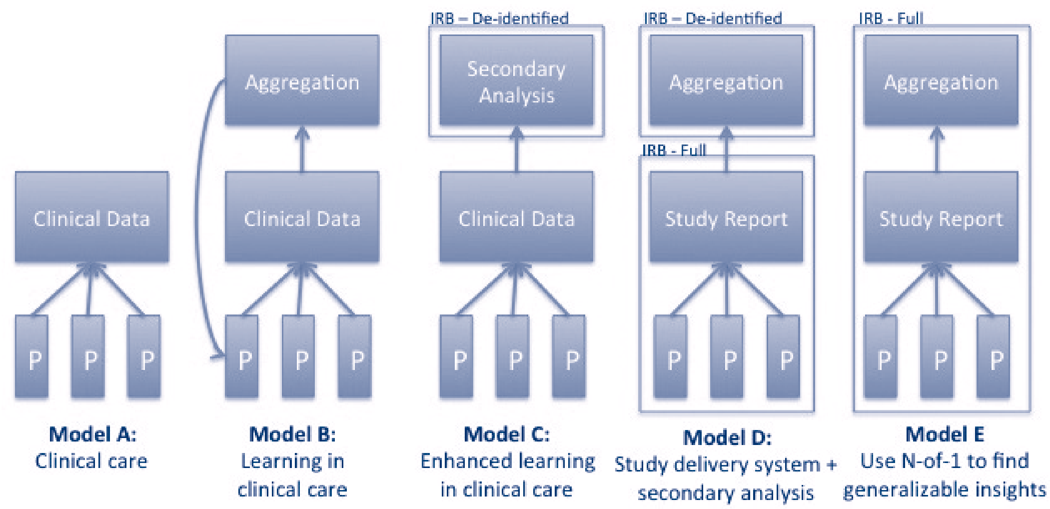
Most importantly, the distinction here lies in the primary objective of the n-of-1 trial. If the primary interest is to produce local knowledge to inform treatment decisions for individual patients, n-of-1 trials so conducted should be interpreted as *clinical care*, and in our view are not subject to the HHS protection of human subjects regulations. Alternatively, if the primary interest is to produce generalizable knowledge to inform treatment decisions for future patients, such n-of-1 trials should be interpreted as *human subjects research* and required to comply with the standards of such research.

Figure 2–1 illustrates five ways in which an organization can conduct n-of-1 trials and use the resulting data. The specific intentions driving the use of a given model of n-of-1 trials inform whether the use of information gleaned from patients should be considered clinical care or human subjects research. We present three case examples to explore the models in greater depth.

Case Examples

Case 1: *The parents of an 8-year old girl diagnosed with attention deficit/hyperactivity disorder (ADHD) come into her physician's office concerned about their child's sleep problems. The physician is aware that increased sleep onset latency is a major side effect of*

Figure 2–1. Five models of n-of-1 trials



Abbreviations: IRB = Institutional Review Board, P = patients/participants.

Table 2–1. N-of-1 trial service compared with research and routine clinical care

Characteristic	Routine Clinical Care[a]	N-of-1 Clinical Service[a,b,c]	N-of-1 Trials Conducted as Research[d,e]
Motivation	Self-interest Intent is to help patient	Self-interest Intent is to help patient	Altruism (greater good) May or may not be helpful to patients
Goal	Optimal patient care (individualized)	Optimal patient care (individualized)	Generalizable data (population estimates of treatment effect)
Population	Based on clinical expertise Consult based Referral based	Based on clinical expertise Consult based Referral based	Inclusion/exclusion criteria Recruit (i.e., advertise)
Informed consent	Yes Procedures, etc.	Yes, n-of-1 approach is a choice NB: Secondary analysis will require separate IRB approval	Yes, participation in research is a choice
Intervention (dose, duration, frequency, route)	Individualized	Individualized	Standardized
Randomization	No	Yes	Yes
Blinding	No	Yes	Yes
Outcomes	Informal	Formal outcomes (will be part of informed consent)	Formal outcomes (data collection)
Publish results	Yes (case reports, series)	Yes (suggest obtain consent a priori)	Yes
Cost of product (discussed further in Chapter 3: Financing)	Varies per jurisdiction	Varies; optimally no charge to patient	No charge to patient
Oversight	Physician licensing board or regulatory college Ensures standard of care	Physician licensing board or regulatory college oversees standard of care; IRB would be involved for secondary analysis	IRB

[a]Corresponds with Figure 2–1, model A
[b]Corresponds with Figure 2–1, model B
[c]Corresponds with Figure 2–1, model C
[d]Corresponds with Figure 2–1, model D
[e]Corresponds with Figure 2–1, model E
Abbreviation: IRB = Institutional Review Board.

stimulant medications. Since there is no approved pharmacologic intervention for sleep problems in children, the physician believes a popular natural health product, melatonin, would benefit this patient. The physician decides to evaluate the effectiveness of melatonin in an n-of-1 trial in which the patient undergoes randomly alternating weeks of 3 mg/day melatonin and identical placebo. Neither the physician nor the parents nor the child will be aware of which treatment the child will be on each week. The parents are asked to monitor their child's sleep and note in a daily sleep diary how long it takes the child to fall asleep over the 6-week period. If the child complains of any side effects throughout the trial, they should notify their physician's office to determine if she should be seen. After 6 weeks, the physician unblinds the random assignments and graphs the results of the trial using the data recorded in the sleep diary. The physician explains the results of the trial to the parents and the child, and together they decide how to proceed. As is, this case exemplifies Model A. If the physician does the same with other patients and draws a general conclusion about whether to continue this approach, this case is an example of Model B. If a researcher later includes this case in a secondary aggregate analysis, it exemplifies Model C.

Case 2 (an example of Model D): *Inflammatory bowel diseases (IBD) such as Crohn's or ulcerative colitis manifest a set of symptoms that are not always correlated to measures of disease activity for which existing therapies have been developed. Patients often try complementary therapies to manage symptoms such as abdominal bloating, urinary urgency, or nighttime stooling patterns. Little is known beyond anecdotes about the efficacy of therapies such as probiotics, dietary manipulation, and herbal medications. A hospital is interested in evaluating whether the introduction of an n-of-1 trial service to IBD clinicians improves the quality-of-life measures for patients who are trying to manage poorly understood symptoms or trying complementary therapies. A researcher at the hospital designs a two-armed study to measure quality-of-life outcomes alongside primary disease activity measures, randomizing clinicians to a control arm (measurement only) and a treatment arm (n-of-1 trial service). The specific n-of-1 study design is determined by the individual clinician-patient dyads, as in Case 1, but the training, data collection procedures, and analysis are submitted to the institution's review board for approval. A secondary analysis can be done separately to assess whether there is evidence of efficacy of a given complementary therapy that may indicate a more structured, parallel group trial of that therapy for symptom management.*

Case 3 (an example of Model E): *Chronic pain is a common condition that has considerable effects on an individual's quality of life. A group of researchers are trying to determine which type of nonsteroidal anti-inflammatory drug (NSAID) will be most effective for chronic pain in adults with osteoarthritis. They decide to conduct a study in which each participant will be enrolled and offered his/her own n-of-1 trial. Each participant will undergo three pairs of 1-week periods of 3,000 mg/day acetaminophen or 1,200 mg/day ibuprofen, for a total duration of 6 weeks. The order of treatments will be randomized for each participant, according to a computer-generated randomization schedule. Patients, doctors, and research assistants will be blinded to treatment order. Participants will be required to mark their pain on a 10-point visual analog scale daily for 6 weeks. At the end of each 6-week trial, participants will receive their individual results. After all n-of-1 trials have been completed, these data will be aggregated to provide an overall estimate of treatment effect.*

N-of-1 Trials as Clinical Care (Model A)

In case example 1, the n-of-1 trial is being used to advance clinical innovation, that is, the patient's health and well-being are of primary interest. Rather than using a novel therapy, the clinician takes a novel approach to assess therapeutic effectiveness (the n-of-1 trial) rather than the usual trial of therapy undertaken by most clinicians. Although randomization, blinding, and use of placebos are unusual in clinical care, their presence alone does not mean the patient's interests are not foremost, as these should be in any clinical encounter.

Learning in Clinical Care (Model B)

The results of n-of-1 trials of clinical care are typically stored in electronic medical records managed by the clinical care provider or by the trial service itself. The availability of electronically accessible data provides opportunities for learning from experience in clinical care, also referred to as evidence farming[8,9,10] or using an evidence macrosystem.[10] Duan[8] characterized evidence farming as a "bottom-up" paradigm for clinical practices to incorporate practice data systematically as a source of evidence, or an articulated form of clinical experience. Hay et al.[9] reported that most physicians participating in a pilot acceptability study saw evidence farming as a promising way to track experience, making scientific evidence more relevant to their own clinical practices. This learning paradigm is especially pertinent for n-of-1 trials, the design and implementation of which are managed primarily in clinical practices.

Learning in clinical care can occur in many ways, with distinct implications for human subjects procedures. In Model B, learning is focused on utilizing the experience from the operations of earlier trials to inform future trial operations, such as the selection of assessment instruments, the number of treatment periods to be tested, etc. This limited learning paradigm does not directly influence clinical care and therefore should be exempt from IRB review.

Enhanced Learning in Clinical Care (Model C)

In Model C, learning requires the major extra step of outcome analyses using de-identified data aggregated from previous n-of-1 trials to inform clinical care decisions in future trials. Here the individual cases being combined are prospectively planned and often randomized and blinded, making them more rigorous in terms of estimates of treatment effect than standard chart reviews of trials of therapy. Since the results of such analyses are used to make decisions about efficacy and not just operations, it is appropriate to seek institutional ethics approval to do secondary analysis for research purposes. These analyses are appropriate for expedited review, like any chart review.

Study Delivery System + Secondary Analysis (Model D)

The more challenging case to be made to an IRB is when individual n-of-1 trials are used as part of a larger intervention on care delivery, for example, studying the impact of an n-of-1 trial service on a hospital or care network. (This is illustrated in Case 2.) In such cases, the entire n-of-1 trial platform should be subject to full ethics review, but the individual trials would not be subject to ethics review, since they would be developed on a clinical basis. Data produced by these trials can also be used for improving clinical care without review and for secondary analysis or meta-analysis with expedited review. If the introduction of a trial service into care is the sole purpose of giving individual clinicians better tools to care for individual patients, and no larger research agenda is addressed, it may be reasonable to assume that no external IRB review is needed.

Use of N-of-1 Trials To Produce Generalizable Insights (Model E)

Finally, it is increasingly common to use a set of identically designed n-of-1 trials to answer questions typically posed in the context of conventional population-based trials, as in Case 3. Analytical techniques to aggregate the data have been developed to facilitate these kinds of trials (see Chapter 4 for details). In these cases, the entire framework of patient recruitment, trial design, data collection, and analysis should be reviewed by the IRB. The critical distinction here is that the design of individual patient trials is dictated by a larger research agenda. Under these circumstances, the autonomy of individual clinicians and patients is limited to ensure that the aggregation of trial outcomes meets the research design goals.

Summary: Role of the IRB Review

Practically speaking, perhaps the most important issue for the implementation of n-of-1 trials is the role of the IRB review. Depending on the primary goal for the trials (research or clinical care), various scenarios are possible:

1. No IRB involvement at all, as the n-of-1 trials are conducted for purely clinical purposes. In this instance, the intervention dose, choice of control, and period length would be individualized to meet the needs of the specific patient. In addition, the use of prior n-of-1 outcomes to improve the design and execution of trials would also be

exempt from regulations for human subjects research.

2. IRB approves platform and procedures of the n-of-1 trials and leaves subsequent treatment selection and design decisionmaking to informed patients and clinicians.
3. IRB approves the n-of-1 trials protocol for a specific condition and specific set of treatments (A, B, etc.). This is appropriate for novel therapies and use of n-of-1 in traditional group trials.
4. IRB reviews and approves each entry into an n-of-1 trial (case by case). This scenario is likely to be prohibitively costly and time consuming (further diminishing the "value proposition" discussed in Chapter 3 on finance), but undoubtedly some will advocate for this. We believe this approach is inconsistent with how research is defined by the U.S. HHS/OHRP. Furthermore, it would create a level of burden that would preclude the use of n-of-1 trials and act in practice to reduce patient choice, which ethically may be considered a kind of harm.

Informed Consent

Informed consent is required from all patients participating in n-of-1 trials. However, the scope of that consent depends on the primary goal of the trial (human subjects research or quality improvement for clinical care).

Equipoise

Equipoise is reached when a rational, informed person has no preference between two (or more) available treatments.[11] While equipoise is usually considered in the aggregate for parallel group RCTs, more specific equipoise on the individual level may be warranted for n-of-1 trials, especially for applications of n-of-1 trials to inform treatment decisions for individual patients. A prerequisite for conducting an n-of-1 trial for an individual patient is that there is substantial uncertainty, given the clinical knowledge available regarding the specific patient, regarding the pros and cons for the treatment options under consideration. More specifically, if there is good reason for the clinician to believe that treatment A is superior to treatment B for the specific patient, it might be unethical to conduct an n-of-1 trial for this specific patient to inform his/her treatment decision. At the same time, such a conundrum should not occur if the informed consent adequately presents the knowledge available, informing the patient of the clinical rationale for preferring treatment A over treatment B.

It is of course possible that the patient, even after receiving careful explanation of the clinical knowledge available, might still have a strong preference for treatment B over treatment A, and requests that an n-of-1 trial be conducted to determine whether the a priori clinical knowledge in favor of treatment A indeed applies to him/her specifically. The clinician could honor the patient's preference in such a situation, as an informed choice by the patient.

Another reason for conducting an n-of-1 trial might be to satisfy a payer regarding treatment effectiveness. In this circumstance, the patient may have a preferred treatment but still be willing to participate in an n-of-1 evaluation so as to gather rigorous data that will allow a payer to be satisfied that the treatment expense is worthwhile.

Publication

Although many IRBs might interpret the intention to publish study findings as a criterion for research being subject to human subjects regulations, it is important to note that the U.S. HHS/OHRP does not necessarily hold this interpretation:

> Planning to publish an account of a quality improvement project does not necessarily mean that the project fits the definition of research; people seek to publish descriptions of nonresearch activities for a variety of reasons, if they believe others may be interested in learning about those activities. (http://answers.hhs.gov/ohrp/questions/7286)

Therefore, it is conceivable that a publication may be derived from a series of n-of-1 trials conducted for the purpose of quality improvement for clinical care, without necessarily subjecting these trials to requirements, such as informed consent for human subjects research, beyond what is required within the realm of clinical care. Whether designed and conducted as research or clinical care, n-of-1 trials may be suitable for publication. Existing trial registries (e.g., clinicaltrials.gov) are compatible for registration of n-of-1 trials, so as to reduce potential for bias from selective publication. Published n-of-1 trial reports should be CONSORT compliant (see CONSORT Extension for N-of-1 Trials).[12]

Summary

In summary, whether n-of-1 trials represent clinical care, quality improvement, or research depends on their intent. If they are designed to improve the care of an individual patient, it is reasonable that they be considered clinical care. Quality improvement can be applied to n-of-1 trials, just as it is in usual care. However, n-of-1 trials may also be considered research if they were designed to answer a larger question for a population of patients. The following checklist was created to help clinicians, investigators, and institutional research ethics boards determine the most appropriate approach in determining which model of n-of-1 trial to use, and what kind of ethics review and approval is required.

Checklist

Guidance	Key Considerations	Check
Clarify intent for n-of-1 trial	• Is the primary reason for designing an n-of-1 trial to improve clinical care for a single patient? Or is it intended to help improve care for the population of patients with that condition? If the primary intent is generalizable data, then the n-of-1 trial should be considered research.	❐
Select model of n-of-1 trial design	• Model A: Clinical care—no IRB approval sought/required. • Model B: Learning in clinical care (analogous to quality improvement)—no IRB approval sought/required. • Model C: Clinical care with secondary analysis—expedited IRB approval sought for secondary analysis of de-identified aggregate data. • Model D: Study delivery system with secondary analysis—full IRB approval sought for the study delivery system (i.e., n-of-1 trial service vs. usual care); expedited IRB approval sought for secondary analysis of de-identified aggregate data. • Model E: Use n-of-1 to find generalizable insights—full IRB approval sought.	❐
Does informed consent need to be obtained?	• Informed consent of patients/participants is needed in all models of n-of-1 trials. • Prospective consent for secondary data analysis is preferred whenever possible.	❐
Equipoise	• There should not be a clinical preference for or against one of the treatments based on health outcomes; however, there can be a difference in preference based on cost or convenience.	❐
Publication	• Whether designed/conducted as research or clinical care, n-of-1 trials may be suitable for publication. • Existing trial registries (e.g., clinicaltrials.gov) are compatible for registration of n-of-1 trials, so as to reduce potential for bias from selective publication. • Published n-of-1 trial reports should be CONSORT compliant (see CONSORT Extension for N-of-1 Trials).	❐

References

1. Guyatt G, Heyting A, Jaeschke R, et al. N-of-1 randomized trials for investigating new drugs. Controlled Clinical Trials 1990;11:88-100.
2. Rothwell PM. External validity of randomized controlled trials: "to whom do the results of this trial apply?" Lancet 2005;365(9453):82-93.
3. Larson EB. N-of-1 clinical trials- a technique for improving medical therapeutics [Specialty Conference]. The Western Journal of Medicine 1990;152:52-56.
4. Guyatt G, Jaeschke R, McGinn T. Therapy and Validity: N-of-1 Randomized Controlled Trials. Users' Guides to the Medical Literature: A manual for Evidence-Based Clinical Practice. Chicago, IL: American Medical Association, 2002: 275-290.
5. Zucker DR, Ruthazer R, Schmid CH. Individual (n-of-1) trials can be combined to give population comparative treatment effect estimates: methodological considerations. Journal of Clinical Epidemiology 2010; 63(12):1312-1323.
6. Kass NE, Faden RR, Goodman SN, et al. "The research-treatment distinction: A problematic approach for determining which activities should have ethical oversight," Ethical Oversight of Learning Health Care systems, Hastings Center Report Special Report 43, no.1 (2013): S4-S15. DOI: 10.1002/hast.133.
7. U.S. Department of Health and Human Services. (2009). Human Subjects Research (45 CRF 46). Washington, DC. Retrieved from www.hhs.gov/ohrp/policy/ohrpregulations.pdf.
8. Duan N. A Quest for evidence beyond evidence-based medicine: Unleashing clinical experience through evidence framing. Presented at UC Davis School of Medicine, October 17, 2002, Sacramento, CA.
9. Hay MC, Weisner TS, Subramanian S, et al. Harnessing experience: Exploring the gap between evidence-based medicine and clinical practice. J Eval Clin Pract. 2008;14(5):707-713.
10. Sim L. Evidence framing: Implications for open arcitechture. Presented at Mobile Health, Stanform University, California, May 5, 2011. Slides available online at www.slideshare.net/openmhealth/evidence-farming-implications-for-open-architecture.
11. Lilford RJ, Jackson J. Equipoise and the ethics of randomization. Journal of the Royal Society of Medicine 1995;88(10):552-559.
12. Shamseer L, Sampson M, Bukutu C, et al. CONSORT extension for N-of-1 trials (CENT) guidelines. BMC Complementary and Alternative Medicine. 2012;12(Suppl 1):410.

Chapter 3. Financing and Economics of Conducting N-of-1 Trials

Wilson D. Pace, M.D.
Elizabeth W. Staton, M.S.T.C.
University of Colorado
American Academy of Family Physicians National Research Network

Eric B. Larson, M.D., M.P.H.
Group Health Research Institute
University of Washington

The DEcIDE Methods Center N-of-1 Guidance Panel*

Introduction

The use of n-of-1 trials to improve therapeutic decisionmaking and clinical outcomes has been studied and reported upon for over 25 years.[1] Selected reports suggest successful resolution of therapeutic uncertainty in specific patients when the underlying condition and drugs are amenable to the n-of-1 approach: specifically, chronic conditions that do not change rapidly over time, with noncurative interventions, clear symptoms that can be tracked, and treatment effects that wash out relatively rapidly (see Chapter 1 of this User's Guide). In a time of increasing interest in personalized medicine, the n-of-1 trial presents a theoretically feasible and cost-effective method of determining the best therapeutic option for a particular person.[2-5] As patients and clinicians recognize that an individual's response to a medication may not be well represented by a population mean, the use of n-of-1 trials to distinguish true responses would seem logical. Nonetheless, after more than 25 years of sporadic reports on n-of-1 trials, largely from academic settings, to our knowledge the service is not generally available to patients and doctors anywhere in the world. This chapter will explore what is understood about costs, benefits, and possible financing of n-of-1 trials based on the literature and the authors' (WDP, EBL) experience.

Although health care providers have access to an array of tools that lend a high degree of confidence to diagnoses, few if any widely available tools help providers determine which medication (or behavioral health treatment) is best for a specific patient. Providers rely on several imperfect strategies for therapeutic decisionmaking. First, they interpret the evidence from randomized controlled trials, which present the average benefits and risks of a particular drug. Such evidence sometimes requires clinicians to find and interpret a large number of studies, then assess the extent to which their patient resembles or differs from the narrow population that qualified for inclusion in the study[6] and the degree to which the benefits and risks of the drug matter to that patient.[7] Second, clinicians may adopt a "trial of therapy" approach, in which they start a patient on a drug and wait to see how it works. The biases and potential problems of this approach have been well described.[7-9] At times, clinicians may simply give a patient two or more drugs in a similar class to take home and try at the patient's convenience (essentially an open-label n-of-1 trial without any control for washout periods, placebo effect, or numbers of crossovers required for clinical decisionmaking). The therapeutic decisions that result from these methods are imperfect at best, and at worst may lead to unnecessary costs and higher than necessary rates of adverse effects.

Beyond improving initial therapeutic decisionmaking, an n-of-1 trial has a number of other potential longer term benefits. In theory, the

*Please see author list in the back of this User's Guide for a full listing of panel members and affiliations.

risk-benefit ratio of a drug would be improved because only medications with demonstrated effectiveness for a particular patient (as shown through an n-of-1 trial) would be prescribed. In addition, short-term side effects are typically clearly demonstrated in n-of-1 trials. Long-term adverse events, of course, are not immediately known and would need to be factored into a risk-benefit model using population-based data. Current population-based information from randomized controlled trials may make it difficult to extrapolate the full benefit of medications in a heterogeneous population. Subgroup analyses can help overcome some of these issues, but studies are often not large enough for these subanalyses, nor are data typically available at the patient level across studies to allow others to examine the heterogeneity issue. When they are feasible, n-of-1 trials eliminate concerns about population-based heterogeneity of responses.

N-of-1 Methods Not Yet Part of Routine Care

Despite their many potential benefits, n-of-1 trials have not become part of mainstream clinical medicine,[10] and to our knowledge have never been a covered benefit in any insurance plan (private or government run) in the United States or Canada. A 2010 systematic review found 108 unique trial protocols from the years 1986 to 2010; the vast majority had authors from Canada (35%), Europe (26%), or the United States (22%).[11] N-of-1 trial services have been run almost exclusively by academic centers with little reach into community practice in the United States; somewhat broader reach has been achieved in Australia.[12] As academically run services, most have been supported by grants and local institutional funds. A systematic review by Gabler et al. found that most trials (69%) reported receiving Institutional Review Board (IRB) approval, and more than half (52%) received external funding.[11] No articles reported charging patients or insurance companies for the service. The involvement of IRBs in the majority of trial activities indicates the low acceptance of these activities as a component of routine clinical care. A more complete discussion of the role of IRBs in n-of-1 trials is presented in Chapter 2. Considering just the impact on financing, the involvement of IRB review for many services (even if the final decision is that n-of-1 trials are not "research") highlights the "experimental" nature of the process and makes insurance coverage less likely, as insurance companies rarely pay for research activities.

Cost Data for N-of-1 Trials

While there is no literature on third-party or patient payments for n-of-1 trials, three studies have explored the costs of conducting trials (see Table 3–1). The reported costs vary widely, partly perhaps because of differences in costs between countries (one article from the United States, one from Canada, and one from Australia) as well as inflationary differences (1993 U.S .dollars versus 2008 Canadian dollars, for instance). Beyond these variables, different approaches have been advocated for conducting n-of-1 trials. Many trial centers develop new trial instruments for each patient, based on the patient's stated preference or importance of one symptom or sign over another. Others report on multiple trials based around a single clinical question, using a standardized set of assessment instruments. Some trials provide feedback to the referring physician, who is then expected to develop a treatment plan with the patient based on the trial results. Other trials include the final treatment decision discussion in the trial itself. These and other differences in approaches would be expected to affect the overall cost of a single trial. In this report, we do not attempt to standardize costs to a particular reference point but simply express costs as reported in the papers we found.

Scuffham et al. evaluated the detailed costs of two multipatient n-of-1 trial series conducted by the University of Queensland.[3] Using classic economic approaches, they initially divided costs into fixed startup costs and variable per-patient or per-trial costs. The costs were considered within the context of a "research" activity conducting two sets of n-of-1 trials using the same medications, the same outcome and side effect instruments, and the same patient problems within each set. The research approach clearly affected the costs incurred and may have also determined whether some costs were considered fixed or variable. The following items were considered fixed costs:

- Seeking funding

Table 3–1. Fixed and variable costs from published n-of-1 trials

Drug	Reference	Country/ Currency	Year of Study	Fixed Costs/ Patient	Variable Costs/ Patient	Cost Diff (n-of-1 Minus Control)/ Patient/ Time
Various	Larson[9]	U.S.	1990	$500	Not reported	Not reported
NSAIDs	Pope[4]	Canada	2002–03	Not reported	Not reported	$31.91/ 6 months
Celecoxib	Scuffham[3]	Australia	2003–05	$1,164	$610	$39/ 12 months
Gabapentin	Scuffham[3]	Australia	2003–05	$1,164	$577	$876/ 12 months

Abbreviation: NSAIDs = nonsteroidal anti-inflammatory drugs.

- Developing the research protocol
- Obtaining ethical review
- Developing instruments/forms for data collection
- Developing treatment sequencing
- Blinding medications
- Design/preparation of medication packs
- Database development

Variable costs were categorized as follows:

- Patient recruitment
- Managing the operation of each trial
- Data collection
- Data analysis
- Generation of results and feedback to clinicians and/or patients

The Queensland trial service found a total fixed cost of $23,280 Australian (2005) to set up two different n-of-1 trial protocols. Various components of these costs would not be applicable when operating n-of-1 trials primarily for clinical purposes. For example, the cost analysis included as "fixed costs" the applications for grant funding and ethical approval, which accounted for $7,730, or 33 percent of the total startup costs. While patient recruitment could also be considered a "research" expense, one would imagine that a commercially available n-of-1 trial system could incur major costs marketing the services to clinicians or patients, which would likely markedly exceed the relatively low "recruitment costs" assigned to this analysis. The cost of preparing medications is listed as a fixed cost, though if medication acquisition costs were included and a broad set of medications were included for potential n-of-1 trials, this would more logically be a variable cost. The investigators considered the developed protocols to be reasonably applied to 200 people, with resultant fixed costs per patient of $116. Variable costs were $610 for a trial of celecoxib versus long-acting acetaminophen and $577 for a trial of gabapentin versus placebo. The overall cost per trial based on this study is in line with many other diagnostic tests. However, this trial did not include costs related to the development of an electronic data collection system, which would be essential for any present-day commercial or clinically based system in the United States or Canada. Even though development of such a system could run into the hundreds of thousands of dollars (U.S.), if the system were used for enough trials, the overall cost per patient could still be kept in line with complex diagnostic tests such as advanced imaging modalities.

N-of-1 trials performed outside of a research study can provide further insight into the costs of the method. One of this chapter's authors (EBL) worked with colleagues to explore the costs of operating an n-of-1 trial service in an academic institution. This service was operated for clinical purposes and therefore did not "recruit" patients as a research protocol would. After initial interactions with the local IRB, the service was declared to be a component of clinical care, therefore not requiring IRB review of each new trial protocol.

Larson's group designed each single-patient trial in their series individually.[5] Their cost assessment then focused on assessing the direct costs of operating a single trial. They estimated 16.75 hours of staff time per trial, which included a physician lead, nursing, data entry, analysis, and feedback time. Of note, none of the staff were solely devoted to work on trials but charged time to the n-of-1 trial service alongside other job tasks. In 1990 U.S. dollars this was estimated at approximately $500/study plus the cost of the medications. In 2013 dollars, just the staff time would likely rise to between $1,500 and $2,000.

Additional experience comes from a commercial application of the n-of-1 model. One of the authors (WDP) worked as an independent evaluator for a commercial venture that sought to bring n-of-1 trials to clinicians in a much more automated form. The group's systems were tested initially with two treatment periods (medication 1 then medication 2, or vice versa) over three treatment cycles.[13,14] This approach was adjusted to five treatment cycles, generally running 5 to 7 days per treatment period, depending on the medication being studied. The group developed a Web-based data collection system and used a validated set of symptom and side-effect questionnaires for the drugs they offered for study. They offered all three primary types of n-of-1 trials: active versus placebo, active drug A versus active drug B, and dose A versus dose B of the same drug. The system allowed clinicians to simply write a prescription for the study of interest from a predetermined set of medications. The company contacted the patient and established a secured Web account. A contracted pharmacy prepared the medication unit dose packs with over-encapsulation to achieve patient blinding. The initial medications available for study were H2 blockers, proton pump inhibitors, and antihistamines. The underlying study design was set by default as requiring five treatment cycles (i.e., AB or BA, where both A and B represent either study medication 1 or study medication 2), with the ordering of treatment periods within each cycle established by random assignment. The number of days per treatment period was determined by the longest half-life of the medications under study, allowing for adequate time to assess symptoms and side effects after a washout period for each medication. If two active comparator drugs were used, patients were crossed over from one active drug to another, without a placebo washout period. To account for the lingering effects of the previous active medication, patient data gathered during a predetermined washout period were ignored in the analysis. A randomly selected crossover pattern was sent to the pharmacy, which prepared the medications for each participant. Analytics were built into the database as a report feature. Clinicians could receive reports as a hard copy or log in to the Web site for the information, including which medication improved symptoms the best, which had lower side-effect profiles, and whether there was a clinically meaningful effect versus placebo.

Unfortunately, the evaluation of the system was stopped early due to financial problems. Prior to that, a total of 64 patients were enrolled; 34 were enrolled in one of two n-of-1 drug trials comparing two active medications using the same data collection system, but only three patients completed full evaluations. This unwelcome experience in the clinical setting differed sharply from the initial, shorter testing, which had very high completion rates.[13] Qualitative feedback indicated that patients did not see enough value in the added certainty provided by the trial results, given that they needed to complete daily logs on symptoms and adverse events for approximately 2 months. Patients indicated they could easily conduct their own open-label trials quickly and inexpensively to determine which medication worked best for them. This finding may have been influenced by the fact that all the medications being studied became available over the counter by the time the evaluation was underway. Interestingly, the side-effect rates (which study data showed were clearly caused by the medication in question, based on study completers or partial completers) were much higher than reported in the literature or package inserts for the medications, approaching 30 percent of proton pump inhibitor users, for instance. This experience is consistent with reports of new or more common significant side effects when drugs are approved by the Food and Drug Administration (FDA) and used in the more general population compared with the highly selected persons typically enrolled in studies meeting FDA efficacy standards.[15]

Cost Offset

To examine cost offsets, Scuffham et al. examined data from two separate multipatient n-of-1 studies conducted in Queensland, Australia: one study compared cox-2 inhibitors versus acetaminophen for osteoarthritis, and the other compared gabapentin versus placebo for neuropathic pain.[3] They constructed a decision analysis model with two arms: “n-of-1 trial” and “no trial.” In both studies, the n-of-1 arms ended up costing more per patient than the “no trial” groups, even taking into account savings from individuals who were able to stop taking ineffective medications. After estimated average per-patient cost offsets of Australian $569 for the gabapentin trial and Australian $221 for the celecoxib trial, the final estimated 5-year additional costs of n-of-1 trial versus no trial for these medications were Australian $869 for gabapentin and Australian $1,152 for celecoxib. This finding could be due to the small differences in outcomes between the n-of-1 and no-trial groups and the low responder rates for both active medications: 17 percent for celecoxib and 24 percent for gabapentin in the n-of-1 trial groups. Both groups demonstrated small improvements in quality of life for the n-of-1 trial participants, resulting in a cost per quality-adjusted life year (QALY) gained in the first 12 months of Australian $36,958 for the gabapentin group and Australian $126,661 for the celecoxib group. If therapy was maintained until the end of life, the cost per QALY gained dropped to Australian $1,725 and Australian $10,278, respectively. The variables most responsible for cost differentials were calculated. These included the underlying variable costs of conducting n-of-1 trials, the number of individuals among whom the fixed costs are shared, the probability that the n-of-1 trial will result in use of the study medication, the time horizon for which the results are valid, and the cost differential of the medications being studied. The longer the patients in this report were credited with taking medication of no value or causing undesirable side effects, the more value would issue from an n-of-1 trial, implying greater cost effectiveness. The paper examined time horizons of 5 years and lifetime, though other studies have used time horizons of less than 1 year following n-of-1 trials.[4,16] The model also indicates that the greater the effect differences between two medications or medication and placebo, the greater the cost efficacy of n-of-1 trials. This analysis used an imputed usual-care group and thus may not entirely capture the impact of an n-of-1 trial at the patient level if a higher percent of people remain on an ineffective drug than imputed.

In examining other reports of multipatient n-of-1 trials (i.e., series of n-of-1 trials entering multiple patients into the same n-of-1 protocol), it is evident why cost offsets can be hard to demonstrate. In the Queensland trials the pain difference for the n-of-1 trial participants versus no-trial group at the end of the celecoxib trial was 0.28 points on a 10-point scale, while the gabapentin trial demonstrated a 0.11-point drop in pain compared to the no-trial group.[17,18] Similarly, in a study of theophylline in patients with chronic obstructive pulmonary disease (COPD), Mahon et al. found that in 68 patients randomized to an n-of-1 trial versus usual care, 7 of the 34 n-of-1 trial patients benefited from theophylline, while 11 elected to continue theophylline at 3 months (35%).[16] By the end of the trial at 12 months, 16 of 34 n-of-1 participants were using theophylline (47%). In the usual-care group, where theophylline effectiveness was determined through open-label on-off usage, 13 of 30 (43%) were using theophylline at 3 months and 15 (50%) were using theophylline at 12 months. Furthermore, there was no difference across study populations (responders and nonresponders in both groups included in the intent-to-treat analysis) in chronic respiratory disease questionnaire scores or 6-minute walk times.

In a study of the use of nonsteroidal anti-inflammatory medications (NSAIDs) in osteoarthritis, Pope et al. found no differences in use of NSAIDs between n-of-1 trial participants and usual-care participants (81% n-of-1 vs. 79% usual care).[4] This relatively small trial (N = 51) found no significant differences in an overall health assessment scale, osteoarthritis pain and function scale, or SF-36 scores between the two groups. The total costs of care (osteoarthritis treatment), including the n-of-1 trial, at 6 months was $551.66 +/- $154.02 for the n-of-1 trial versus $395.62 +/- $226.87 for the usual-care group (2003 Canadian dollars). Since n-of-1 trials, even if taken to scale, will always cost more than open-label clinical trials, it will require demonstrations of greater effect from the trials themselves to demonstrate reasonable cost offsets.

Value Proposition

In general our review concludes that it is difficult to demonstrate a value proposition for n-of-1 trials based on the current literature. Trials reported to date have found limited differences in outcomes between n-of-1 participants and usual care, a tendency of both groups to end up with similar medication usage patterns over time, and small sample sizes. Kravitz et al.[10] have postulated the potential for greater value where treatment costs are higher, such as with biological agents. Furthermore, where risk-benefit equations are very different between various treatments (e.g., low-dose methotrexate vs. biological agents for rheumatoid arthritis), demonstrating clear benefits to higher risk medications may improve the overall value proposition as the population of medication users is enlarged and serious side effects from high-risk medications appear. These issues are not considered in any of the current literature which directly examines costs of n-of-1 trials; given the small sample sizes and short followup timeframes, major side effects from medications were not encountered. A more general issue is that chronic disease effects and available treatments change over time. These changes may lower the enduring value of the results of an individual treatment, given changes in symptom patterns or a patient's preference or physician recommendation based on availability of different treatments. For n-of-1 trials to be valuable in the face of seemingly inevitable changes, the methods would need to be relatively straightforward and efficient and meet patients' timeliness expectations.

Karnon et al. have explored the use of n-of-1 trials to study the economic impact of various medication choices at the individual patient level.[19] The authors consider adding questions related to total cost of care, cost of alternative medications used, and/or quality of life to better understand the cost/benefit of various medication choices. The paper considers the ethical issues of basing decisions on overall improvement versus the cost per unit of improvement. It concludes that clear patient preferences should drive clinical decisions and that economic considerations should come into play only when the clinical decision is ambiguous. The use of a series of n-of-1 trials with additional data collection could help researchers more precisely understand the economic and quality-of-life impact of various medication choices in responders. This rationale could also arguably be applied to diagnostic tests, which are typically adopted and paid for without a clear demonstration of a value proposition other than improved diagnostic accuracy.

Influence of Personalized Medicine

Personalized medicine is an area in which n-of-1 trials may help us study outcomes for commonly prescribed drugs. With growing concern about the overall safety and risk-benefit profile of many medications, n-of-1 trials could be used to personalize this information. N-of-1 trials seem particularly well suited to understanding side effects associated with a medication at the personal level. Could this drive interest in the method, if it were better understood? Similarly, n-of-1 trials are well suited to study herbal preparations, dietary supplements, and behavioral treatments (including lifestyle, behavioral, and complementary/alternative interventions, as discussed in Chapter 2 of this User's Guide). There are many "natural" supplements available for a wide variety of conditions, most of which will never be submitted to rigorous population-level randomized controlled trials. Through crowd-sourcing, could a subgroup of individuals interested in trying supplements form a grassroots user group interested in the therapeutic precision of n-of-1 trials? The Patient-Centered Outcomes Research Institute[20] is developing Patient Powered Research Networks that could form a basis for a patient-centered n-of-1 trial network.

As we move toward personalized medicine based on genomic or proteomic data, combining n-of-1 trials for appropriate conditions and medications may be the one rational way to study outcomes associated with commonly prescribed drugs for both individuals (personalized medicine) and general populations. We can assume that the attractiveness of personalized medicine will grow, and as science-based personalized medicine disseminates, n-of-1 trials seem elegantly suited to become a regular part of personalized medicine.

Potential Financing Options

We have identified a number of potential ways in which the n-of-1 trial could be paid for. It is conceivable that large pharmacy chains could take on the conduct of n-of-1 trials. Most of these companies already have a strong Internet and mobile presence, the ability to prepare the medications for trials, and established financial relationships with payers.

If n-of-1 trials demonstrated positive financial offsets for selected medications, or greater levels of patient satisfaction and improved outcomes with low marginal costs, would Accountable Care Organizations (ACOs) consider contracting with commercial vendors or pharmacy chains for the service for selected medications? It is conceivable that with ACOs and cost bundling, n-of-1 trials would have a value proposition as a strategy to manage expenses while reducing side effects and adverse effects of drugs, especially for commonly used or expensive drugs. The Centers for Medicare & Medicaid Services (CMS) or the new Innovation Center within CMS could be a source for funding that would help elucidate the impact of n-of-1 trials taken to scale in usual clinical care. N-of-1 trials could be considered cognitive services, which do not involve a capital investment in a machine, device, provision, or procedure and thus have little potential in a fee-for-service world to cover implementation or facility costs.

No self-interested group has yet been inclined to develop a business case for n-of-1 trials. If anything, pharmaceutical companies have previously had a disincentive in a fee-for-service world to consider n-of-1 trials, since they typically reduce overtreatment and highlight side effects. In an ACO world, n-of-1 trials could be part of a risk-mitigation strategy to reduce overtreatment and medication side effects. Given the precision of information on short-term side effects developed through n-of-1 trials and the current FDA priority to find better ways of detecting adverse effects post marketing, the FDA might consider whether developing an infrastructure for an n-of-1 enterprise might be a worthwhile way to improve assessment of medications for symptomatic treatment of chronic diseases. If a trial registry were available that contained both standardized methods and outcome assessment toolkits as well as a repository for trial results, data derived from potentially thousands of individually conducted n-of-1 trials could be an added source of information to assess benefits and risks of drugs. Patients in a clinical trial registry would likely represent a broader population in regard to age and secondary morbidities than typically seen in phase 3 randomized controlled efficacy trials, allowing a better understanding of the impact of medications in everyday practice through secondary analysis of the pooled results. We believe this idea is worth exploring, especially for medications that are likely to be taken long term for chronic conditions.

Innovations That May Increase the Appeal of the N-of-1 Trials

Several innovations could increase the reach and appeal of n-of-1 trials. Interactive technology (discussed in Chapter 5 of this User's Guide, which covers information technology) could incorporate patient preferences for the most important outcomes to them (a potentially variable cost) while still maintaining a "standard" data collection format. Validated instruments, for both outcomes and adverse effects, could be built into the data collection system, with patients indicating the most important personal outcome as well as the side effects they consider least tolerable or most troublesome. While the initial costs of development could be substantial, the per-trial cost could still be reasonable if amortized over thousands of patients. However, overall usage of n-of-1 trials would need to expand greatly for this model to be cost effective.

As mentioned, a national n-of-1 trial registry could improve shared decisionmaking based on an individual n-of-1 trial over time. Such a registry would store and analyze the combined results of n-of-1 trials using standardized processes and include well-developed assessment methods and outcome scales for appropriate medications, particularly those with narrow therapeutic windows, moderate population-level efficacy, or high cost-to-benefit ratios. This advance would increase the reach and use of n-of-1 trials greatly (but would likely require substantial ongoing support).

Conclusion

The long-term financing of n-of-1 trials will be determined by a value proposition that is more attractive to patients, clinicians, and other providers, including perhaps pharmaceutical companies, payers, and possibly regulators. Presently the limited use of n-of-1 trials may reflect that the value proposition for clinicians and patients lies with the rapid acquisition of data to guide diagnosis and treatment. N-of-1 trials, with their prolonged timeframe, are relatively unattractive compared to other clinical activities that produce rapid results, even though n-of 1 trials could fundamentally change the way that medicine is practiced. Can n-of-1 trials become more standardized, more efficient, and more patient and physician friendly? Most importantly, can they be moved from the rarefied world of the academic medical center and faculty with keen interests in clinical epidemiology and research to the everyday world of clinical practice and the rapidly changing world of consolidating delivery systems?

Larger scale, more efficient services aimed at enhancing patient-centered outcomes through more precise therapeutics could be a way to demonstrate value. The outcomes of greatest interest would be improved effectiveness of treatment, reduced side effects, and improved patient and physician satisfaction, along with reduced or improved management of costs through avoidance of adverse events and an ability to use less expensive drugs of proven effectiveness for individual patients. For n-of-1 trials to reach a broader audience, it will be important to develop methods that reduce patient reporting burdens. The use of small, Internet-connected personal devices should make this a possibility.

As with diagnostic interventions, an understanding of the characteristics of the intervention is important in determining when it will benefit patients and when it is contraindicated. For diagnostic interventions, these characteristics include specificity, sensitivity, prior probabilities, and positive and negative predictive values. For n-of-1 trials, a better understanding of the impact of different characteristics of the treatment differentials would help advance the concept. For instance, what are the impacts of different probabilities of a positive response to treatment on the utility of an n-of-1 trial? At what level of population response is an n-of-1 trial no longer indicated? What are the impacts of various levels of cost differentials of the final treatments on the potential benefits of an n-of-1 trial? Clinicians need information that will help them understand where n-of-1 trials would be of greatest value.

Overall, we conclude that the limited data currently available suggest that n-of-1 trials can be conducted for a reasonable per-patient cost (not considering the cost of the drug or drugs to be tested) and that these costs could be further lowered with modern technology such as interactive data collection systems. Furthermore, modern technology should be able to blend standardized data collection instruments with patient preference and modern testing theory to reduce data collection from nonuseful questions for a particular patient. The value proposition, from both the financial and patient outcome perspectives, is where the most uncertainty exists at present. Until this value proposition is better defined, it is unlikely that commercial payers will include coverage for n-of-1 trial activities.

Checklist

Guidance	Key Considerations	Check
Consider the cost related to assessment instruments	• Developing new instruments for each patient/trial increases costs. • Standardized assessments reduce analytic efforts later.	❐
Provide feedback	• Feedback to clinicians will help them develop treatment plans. • Feedback can be incorporated into the trial itself.	❐
Plan for fixed start-up costs	• Fixed costs include developing instruments/forms for data collection, developing treatment sequencing plans, blinding medications, designing and preparing medication packs, developing a database, marketing the trials.	❐
Think about additional costs if your service will be considered "research"	• Research costs include seeking funding, completing IRB process, more complicated consent.	❐
Plan for variable per-patient or per-trial costs	• Variable costs include recruiting patients, managing the operation of each trial, collecting data, analyzing data, generating results, and feedback to clinicians and/or patients.	❐
If considering the cost offset, consider relevant elements	• The greater the effect differences between two medications or medication and placebo, the greater the cost effectiveness. • The longer patients take medication of no value or medication that causes undesirable side effects, the more value would issue from an n-of-1 trial, implying greater value and thus cost effectiveness.	❐

References

1. Mahon J, Laupacis A, Donner A, et al. Randomised study of n of 1 trials versus standard practice. BMJ. 1996 Apr 27. 1996;312(7038):1069-1074.
2. Kravitz RL, Duan N, White RH. N-of-1 trials of expensive biological therapies: a third way? Arch Intern Med. May 26 2008;168(10):1030-1033.
3. Scuffham PA, Yelland MJ, Nikles J, et al. Are N-of-1 trials an economically viable option to improve access to selected high cost medications? The Australian experience. Value Health. Jan-Feb 2008;11(1):97-109.
4. Pope JE, Prashker M, Anderson J. The efficacy and cost effectiveness of N of 1 studies with diclofenac compared to standard treatment with nonsteroidal antiinflammatory drugs in osteoarthritis. J Rheumatol. Jan 2004;31(1):140-149.
5. Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a 2-year period. JAMA. 1993 Dec 8. 1993;270(22):2708-2712.
6. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" Lancet. Jan 1-7 2005;365(9453):82-93.
7. Ross JM. Commentary on applying the results of trials and systematic reviews to individual patients. ACP Journal Club. Nov-Dec 1998;129(3):A17.
8. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA. 1994 1994;271(9):703-707.
9. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA. 1994 1994;271(5):389-391.
10. Kravitz RL, Duan N, Niedzinski EJ, et al. What Ever Happened to N-of-1 Trials? Insiders' Perspectives and a Look to the Future. Milbank Quarterly. 2008;86(4):533-555.
11. Gabler NB, Duan N, Vohra S, et al. N-of-1 trials in the medical literature: a systematic review. Medical Care. Aug 2011;49(8):761-768.
12. Nikles CJ, Mitchell GK, Del Mar CB, et al. An n-of-1 trial service in clinical practice: Testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. Pediatrics. 2006 Jun. 2006;117(6):2040-2046.
13. Reitberg DP, Del Rio E, Weiss SL, et al. Single-patient drug trial methodology for allergic rhinitis. Annals of Pharmacotherapy. 2002 Sep. 2002;36(9):1366-1374.
14. Wolfe B, Del Rio E, Weiss SL, et al. Validation of a single-patient drug trial methodology for personalized management of gastroesophageal reflux disease. Journal of Managed Care Pharmacy. Nov-Dec 2002;8(6):459-468.
15. Greener M. Drug safety on trial. Last year's withdrawal of the anti-arthritis drug Vioxx triggered a debate about how to better monitor drug safety even after approval. EMBO Reports. Mar 2005;6(3):202-204.
16. Mahon JL, Laupacis A, Hodder RV, et al. Theophylline for irreversible chronic airflow limitation: a randomized study comparing n of 1 trials to standard practice. Chest. Jan 1999;115(1):38-48.
17. Yelland MJ, Nikles CJ, McNairn N, et al. Celecoxib compared with sustained-release paracetamol for osteoarthritis: a series of n-of-1 trials. Rheumatology. Jan 2007;46(1):135-140.
18. Yelland MJ, Nikles CJ, McNairn N. Single patient trials of gabapentin for chronic neuropathic pain. World Congress on Pain; 2005; Sydney, Australia.
19. Karnon J, Qizilbash N. Economic evaluation alongside n-of-1 trials: getting closer to the margin. Health Econ. Jan 2001;10(1):79-82.
20. Patient-Centered Outcomes Research Institute. 2013; www.pcori.org. Accessed July 8, 2013.

Chapter 4. Statistical Design and Analytic Considerations for N-of-1 Trials

Christopher H. Schmid, Ph.D.
Brown University

Naihua Duan, Ph.D.
Columbia University

The DEcIDE Methods Center N-of-1 Guidance Panel*

Introduction

In this chapter, we discuss key statistical issues for n-of-1 trials—trials of one patient treated multiple times with two or more treatments, usually in a randomized order, with the design under the control of the patient and his or her clinician. The issues discussed include special features of experimental design, data collection strategies, and statistical analysis. For simplicity, we will focus on the two-treatment, block pair design in which patients receive each of two treatments in every consecutive pair of periods with separate treatment assignments within each block of two periods, either randomized or in a systematic, balanced design. Extensions are straightforward to other designs such as K treatments (K > 2) assigned in blocks of size K, randomization schemes with differently sized blocks (e.g., block sizes equal to a multiple of the number of treatments), or unblocked assignment schemes, requiring no changes in the fundamental principles we outline. The basic design principles include randomization and counterbalancing, replication and blocking, the number of crossovers needed to optimize statistical power, and the choice of outcomes of interest to the patient and clinician. Analyses must contend with the scale of the outcomes (continuous, categorical, or count data), changes over time independent of treatment, carryover of treatment effects from one period into the next, (auto)correlation of measurements, premature end-of-treatment periods, and modes of inference (Bayesian or frequentist). All of these complexities exist within an experimental environment that is not nearly as carefully regulated as the usual randomized clinical trials and so require an appreciation of the special difficulties of gathering data in an n-of-1 trial.

Experimental Design

One of the appealing features for the n-of-1 trial lies in its allowing the patient and clinician to devise an individualized trial with idiosyncratic treatments and outcomes run in real-world settings. As a result, n-of-1 designs may vary substantially and reflect great creativity. On the other hand, they often involve clinicians who are unfamiliar with the principles and practice of clinical trials and who may not have access to the resources common in research settings. Because many n-of-1 trials will be carried out in nonresearch medical office or outpatient clinic environments, it is important to ensure that proper experimental standards are maintained while allowing designs to remain flexible and easy to implement. One way to ensure such standards is to establish a centralized service responsible for crucial study tasks such as providing properly randomized or balanced/counterbalanced treatment sequences to the patient-clinician pair when they are designing the trial. We next discuss common clinical crossover trial standards that continue to apply in n-of-1 studies.

Randomization/Counterbalancing

After choosing the identity and duration of the treatments to be given, the patient and his/her clinician must be given a sequence of treatments in such a way that the validity of the experimental

*Please see author list in the back of this User's Guide for a full listing of panel members and affiliations.

process is maintained. The sequence can be either randomized or generated in a systematic counterbalanced design, such as ABBA.[1,2] In the standard two-treatment n-of-1 trial, the assignments are made within blocks of two time periods. With randomization, the first time period in each block is assigned randomly to one of the two treatments, say, A; the second time period is then assigned to the other treatment, say, B. With a counterbalanced design, the assignments alternate between AB and BA in a systematic manner that is intended to minimize possible confounding with time trend. For example, each two blocks can be assigned as AB (first block) BA (second block) to eliminate possible confounding with a linear time trend.

An important requirement for a good experimental design is to balance treatment assignments, especially for potential confounding factors, so that the treatments are compared fairly. Making assignments in blocks of size two ensures that each patient receives each treatment with the same frequency at a comparable set of times, to avoid poorly balanced designs such as AABA and AABB.

Randomization and counterbalancing attempt to balance treatments both within and across blocks. Randomization achieves balance, on expectation, when averaged across a large number of blocks and/or a large number of n-of-1 trials. However, for each individual n-of-1 trial, exact balance might not be achieved. For example, if patient outcomes are deteriorating gradually over time, inducing a time trend, the ABAB design would not be well balanced, as B is always delivered after A. The design itself may induce inferior outcomes for B due to the time trend when the two treatments are actually equivalent. For a four-period trial randomized in blocks of size two, there is a 50-percent chance that randomization will yield such an unbalanced design, either ABAB or BABA (and a 50% chance of a design that is well balanced against the linear time trend, either ABBA or BAAB). Counterbalancing, on the other hand, can be more effective at achieving exact or nearly exact balance for the potential confounding factor(s) designed explicitly to be balanced, for example, the ABBA design achieves exact balance for linear time trend.

While randomization can be less effective than counterbalancing in distributing known confounding factor(s) in a balanced way across treatment periods, randomization has an important advantage in its ability to balance (on average) all potential confounding factors, both known and unknown. Counterbalancing, on the other hand, can perform poorly if the explicit scheme chosen leads to imbalance with respect to an unknown confounding factor.

In addition to reducing but perhaps not completely eliminating the risk of bias induced by time trends, blocked assignment also provides two other important benefits. It minimizes the consequences of early termination from the trial that might otherwise lead to an unbalanced number of observations in the two treatment arms. Within-block assignment also reduces the chances that unknown confounders may bias the estimate of within-patient variation, which would invalidate appropriate statistical inference.

To summarize, we recommend that a blocked scheme for treatment assignment be used for n-of-1 trials. We also recommend that users make a careful choice between randomization and counterbalancing. If there is good information on the most important potential confounding factor (such as the linear time trend) and if the total number of blocks is small, say, less than four, counterbalancing can be more effective. Otherwise, randomization would be a more robust choice. The end of the next section, Blinding, has some further discussion.

Blinding

To the extent possible, patients and clinicians should remain blinded to the treatment assigned, particularly when patient-reported or other subjectively ascertained outcomes are used. While blinding is desirable in all clinical trials, it may be particularly important with n-of-1 trials because of the individualized crossover nature of the study. Patients may (and probably will) try to guess which treatment they received in each period. Because they are so invested in the research and so desirous of a positive outcome, it is natural that their reported outcome measures are affected by knowledge of the treatment received—for example, in favor of the direction that confirms any preexisting expectations they might have (the expectancy effect).[3,4] Potential bias might also ensue from the motivation for the trial if, for example, patients were compelled to enter an n-of-1 trial to prove that a more expensive treatment was really indicated and should be reimbursed. On the other hand, patients' self-interest might also drive them to report as objectively as possible, particularly if they enter the trial without any preconceived preferences, because they themselves will bear the consequences of a bad treatment decision based on biased outcome reports.

In the absence of blinding, other features related to treatment administration might influence outcomes, but in such a way that they should actually be incorporated into the treatment decision, if it is reasonable to expect the same effect will persist beyond the end of the trial. It was noted in the section "Blinding" in Chapter 1: "Patients and clinicians participating in n-of-1 trials are likely interested in the net benefits of treatment overall, including both specific and nonspecific effects." For example, if the patient prefers one pill to the other because of its color or texture during the trial (a nonspecific effect), and this effect is sustained, it is a real effect for this patient and should be part of the treatment decision. In a parallel group trial where the intent is to generalize beyond the patients in the trial, such a preference should be considered a bias, because future patients to be treated according to the findings from the trial might not have the same preference.

In addition to the potential effect on reported outcomes, knowledge of treatment identity may lead some to end a treatment period early if the measured outcomes support the treatment expectation. Even if the treatment assignment is blinded, superior results in one or more periods may induce patients to ask to unblind the trial to confirm whether their hunches are correct. Such unblinding will stop the trial and may result in an inconclusive result.

For blinded n-of-1 trials with treatments assigned in small blocks such as blocks of size two, there is sometimes a concern that some users (patients and/or clinicians) might learn during the course of the trial that the second treatment in the block is predetermined by the first; therefore, the outcome for the second treatment might be affected by expectancy. When this is an important concern, one could use a block size that is a multiple of the number of treatments or randomize the block sizes in different multiples of the number of treatments. This strategy minimizes the chance for the user to figure out the treatment in any given period. On the other hand, this strategy may also increase the risk of bias if time trends are present or dropout occurs.

Replication

Because only one patient is involved in an n-of-1 trial, the number of measurements taken on each individual determines the sample size of the study. The total number of measurements is determined, in turn, by two components: the number of periods and the number of measurements per period. For instance, a pain outcome measured daily over six 14-day treatment periods will have 84 observed data points. These repeated measurements enable estimation of between- and within-period variances, both crucial for proper statistical modeling. Larger sample sizes can be achieved by increasing the number of treatment periods, increasing the length of each period, or increasing the frequency of measurements within each period. These alternative strategies have different analytic implications because they affect different components of the study variance. It is important to carefully choose both the number of crossover periods and the number of measurements taken per period to enhance the efficiency of the study design. More data will improve the precision of the treatment effect estimate, but the optimal allocation to more treatment periods or more measurements per period depends upon statistical considerations such as the expected size of each variance component and its influence on the precision of the effect of interest and the minimum effect size of interest, as well as on practical considerations related to feasibility and type of measurement. Such considerations include patients' ability to record data more than once a day, the validity of measures on different time scales, increased likelihood of dropout with longer trials, and the tendency for patients to become less careful in following treatment protocols over time. Outcomes with substantial measurement variation such as quality-of-life measures will need to be collected more frequently in order to precisely estimate the variance.

Washout

Carryover, the tendency for treatment effects to linger beyond the crossover (when one treatment is stopped and the next one started), threatens the validity of the comparison between treatments in crossover studies, including n-of-1 trials. While statistical models may attempt to accommodate carryover, they rely on assumptions about the nature of the carryover that may be difficult to test or even control. In the extreme, carryover may extend throughout all or most of the next treatment period, contaminating many of the outcome measurements.

Inserting a washout period in which no treatment is given between consecutive treatment periods is the most common method to reduce or even eliminate the effect of carryover by design. The goal of a washout period is to provide time for each patient to return to the baseline disease state, unaffected by preceding treatment. Deciding whether to include a washout period depends on both clinical judgment about the durability of the treatment effect (e.g., from the pharmacokinetics of a drug treatment) as well as practical and ethical considerations related to the study design's implications on the satisfaction and welfare among end-users (patient and clinician).

An important clinical consideration for the washout is to avoid adverse interaction between the treatment conditions. This is mainly an issue for active-control studies, with an active treatment (the standard treatment) used as the control condition to evaluate the comparative effectiveness of an alternative treatment. If the two active treatments being compared are not compatible with each other, it would be necessary to impose a washout period to eliminate the first agent before starting the second agent.

When adverse interaction can be ruled out, the inclusion of a washout period can be problematic for active-control studies, both in terms of satisfaction for the end-users (patient and clinician) and in terms of clinical ethics. The washout period introduces a third treatment condition: the absence of either active treatment. Even a patient managing the disease condition adequately with current treatment might undertake the n-of-1 trial to test the possibility that the alternative treatment might be better. It is undesirable, and perhaps even unethical, for the patient to be forced into a period of no treatment that is likely to be inferior to the current treatment. The use of washout in such studies might reduce a patient's willingness to undertake the n-of-1 trial and increase the chances of early termination from the trial. The ethical dilemma here is that, when adverse interaction can be ruled out, there is no obvious clinical rationale to withhold both active treatments from the patient during the washout period, other than to make a short-term sacrifice in exchange for a better chance to improve the therapeutic precision at the end of the trial.

Conversely, not using a washout might compromise the validity of the estimated treatment effect and lead to biased estimates for treatment effects. Therefore users need to determine whether the likelihood of a substantial bias warrants the drawbacks of the washout.

In some cases, the effect of the washout can be accomplished analytically without including any period during which treatments are withheld. More specifically, any effect of carryover can be dealt with analytically by eliminating, discarding, or downweighting observations taken at the beginning of a new treatment period. It is also possible to incorporate all observations by introducing a smooth transient function that drifts toward zero gradually over time and reflects the time to respond to the carryover effect. Such a function would reduce the influence of potentially contaminated observations early in the period. It contrasts with discrete functions that either accept or discard early observations. This approach can also help to maintain the integrity of the trial by reducing the chance that the patient will drop out and that observations will be contaminated by carryover.

While carryover affects how the effects of the previous treatment might linger after the completion of the previous treatment period, another important transition issue is the onset of the new treatment. Some treatments, such as selective serotonin reuptake inhibitors (SSRIs), may take some time to reach full effectiveness. Slow onset provides another reason to reduce the influence for potentially contaminated data at the beginning of a period; it introduces a natural washout, particularly if the time for one drug to wear off is no greater than the time for the next drug to take effect.

It should be noted that a washout period does not directly mitigate the problem of slow onset. On the contrary, a washout period further extends the transition between the two treatments, because the onset for the new treatment does not begin until the end of the washout period. As an example, assume that treatment A takes 3 days to wash out, and treatment B takes 2 days to reach its full effectiveness. If a washout period of 3 days is used after a period of treatment A, then treatment B begins on day 4 and reaches its full effectiveness on day 6. Therefore, a total of 5 days are lost to the transition between the two treatments. On the other hand, if a washout period is not used (under the assumption that there is no adverse interaction between the two treatments), the transition is 3 days only: by day 3, treatment B has reached its full effectiveness; by day 4, the carryover effect for treatment A has disappeared. Therefore only 3, instead of 5, days of treatment do not reflect full treatment effects.

If a washout period is included in the study design, its length needs to be chosen carefully, taking into consideration treatment interactions, medical ethics, drug half-lives, and onset efficacy. Longer washout periods decrease the likelihood of carryover but increase the length of the study and time spent off treatment, and also delay the onset of the full effectiveness of the next treatment. Making washout periods too short contaminates treatment effects and carryover effects, and might result in biased estimates for treatment effects. In summary, one needs to define treatment periods sufficiently long to manifest the intended treatment effect and overcome transient effects such as carryover and onset, but short enough to allow enough crossovers within a reasonable total duration for the study.

Adaptation

While a fixed trial design is the norm, adaptive trial designs offer the chance to modify the design of an ongoing trial in order to make it more efficient or to fix problems that may have arisen.[5] Some adaptations occur naturally, as when a patient and clinician decide to stop a trial because one treatment appears to be more effective or end a treatment period early because of an adverse event. It is important in such circumstances that blinding be maintained if it is already part of the study design. For instance, it would not be proper to unblind a treatment period in order to stop one treatment, but not the other. Other adaptations could include extending the length of the trial to more treatment periods if treatment differences appear to be small or instigating play-the-winner designs,[6,7] in which the treatment that appears to be more effective is given more frequently. Such designs are generally easier to implement when the data are analyzed using Bayesian methods without tests of hypothesis whose properties depend on prespecified design plans. If frequentist inference (i.e., p-values) is used, sequential design with explicit stopping rules is necessary to protect the overall type I error rate. In some cases, decisions to adapt a design may arise from experience with similar patients. For the implementation of adaptive and sequential designs, it is important that these procedures be built into the informatics system to allow for automation of these design features. In order to ensure high-quality performance of the automated procedure, we recommend that these procedures should be reviewed periodically and calibrated as needed.

Multiple Outcomes

The personalized nature of n-of-1 trials and their focus on making a treatment decision for an individual patient require outcomes to be carefully chosen so as to reflect the measures of most importance to the patient's well-being. Often, more than one outcome is of interest to the patient—perhaps obtaining relief from pain and sleeping better—and so the effect of treatment on both needs to be considered in the choice of treatment at the end of the trial. This contrasts with most clinical trials, which often focus on one particular average treatment effect in the population. Thus, although almost all clinical trials collect data on at least several, if not many, outcomes of interest, they typically focus on a primary outcome and so use statistical methods for a single outcome variable.

A common technique when multiple outcomes are of interest is to form a composite variable such as MACE in cardiovascular trials, which counts the number of major adverse cardiac events (e.g., acute myocardial infarction, ischemic stroke, coronary arterial occlusion, and death), and then analyze it by univariate methods. Composite outcomes are not as popular in n-of-1 studies because they do not allow the patient or clinician to see the effect on each distinct outcome separately. Often the outcomes differ so fundamentally that forming a composite becomes difficult. Returning to a previous example, how might one combine a pain scale and the number of nights of good sleep over a fortnight? One could express both as a percentage of relief compared to a baseline level and then average the two percentages, but this would assume that both outcomes were of equal importance and that both outcome scales were linear. Alternatively, one could choose one outcome as primary and the other as secondary, but if the patient were concerned with both, this would be unlikely to work well. Another approach would be to form a weighted composite scale, with weights accommodating patient and clinician preferences or utilities.

To reflect the patient's true decisionmaking state, one might instead analyze each outcome separately and report a measure of the treatment's effectiveness for each, letting the patient and clinician weight them on their own. One could argue, however, that explicitly specifying the weights up front is more scientific and transparent than having the patient and clinician implicitly weighting separate outcomes post hoc in trying to make a treatment decision. In the end, this is a decision problem, and it is worth exploring methods of decision analysis to improve decisionmaking for n-of-1 trials. Both approaches may be useful.

Because the focus is on the immediate decision of which treatment to take, it is not important to protect against a false-positive decision, as in the standard test of hypotheses commonly employed in clinical trials. One is not choosing to report a statistically significant finding for one outcome among many, so multiple testing is not an issue. Instead, one provides the decisionmaker with all the information required in a format that facilitates decisionmaking.

Multiple Subjects Designs

Several publications have described an n-of-1 service in which many patients are offered the opportunity of carrying out studies. Such services offer several advantages: economies of scale in research infrastructure, clinicians experienced in n-of-1 trials, and the chance to use information gained from other patients. Multiple n-of-1 trials may be combined in a common statistical model to both estimate the average treatment effect as well as improve individual treatment-effect estimates by borrowing strength from the information provided by other similar patients. As more patients accrue, not only does the precision with which the next patient can be evaluated improve, but the estimates for previous patients who might have even finished their studies may also change as a result of information gathered from later patients. Multiple-subject designs increase the complexity of sample size choices, because they permit manipulation of the number of subjects as well as the number of measurements on each. Balancing these two numbers requires knowledge of the relevant within- and between-patient variances.[8] Ethical considerations may also arise from multiple n-of-1 trials if one treatment appears to be working better and clinicians become reluctant to continue randomizing patients due to lack of equipoise.

Data Collection

The lack of research infrastructure for the single clinician running an n-of-1 trial may have a serious detrimental effect on data collection. Typically, research studies initiate elaborate procedures to ensure that data are collected in a timely, efficient, accurate fashion. Forms are tested and standardized; research assistants are hired and trained to help collect data from patients either at patient visits or remotely via mail, telephone, or Internet connections; data are checked and rechecked by trial personnel and external monitors; and missing items are followed up. Many of these options are not available to the typical clinician running a trial outside of an established n-of-1 service. Conversely, patients in n-of-1 trials are usually extremely motivated, because the trial is being done for them and by them, so they may be more committed to data collection and therefore less likely to miss visits and fail to complete forms accurately. Missing items can be particularly costly in an n-of-1 study because of the small number of observations.

Clinicians undertaking n-of-1 trials must be aware that each trial is unique, with its own protocol and its own set of outcomes. This multiplicity of designs can complicate data collection, even if a centralized support service is available. Multiple data collection forms may be needed, and personalized user interfaces may be valuable ways to collect data. Reminders are important to provide, and interim feedback can maintain the patient's enthusiasm.

Statistical Models and Analytics

The unique design features of n-of-1 trials, including a multiple-period crossover design, multiple patient-selected outcomes, and focus on individual treatment effects, motivate statistical models for these trials. Data resemble a time series in that they are autocorrelated measurements on a single experimental unit. Unlike classic time series, however, the measurements are structured by the randomized design, and so statistical models also have features like those for longitudinal data with a time-varying covariate (the treatment condition). The main goal is to compare the observations made under the two treatment conditions, adjusting for any carryover effects, while accommodating the randomized block structure.

Constructing such models is difficult, especially when few measurements are taken. One review of the n-of-1 literature in medicine, in fact, found that many studies used no formal statistical model at all to compare treatments, opting instead for eyeball tests based on a graph of the data or simple nonparametric tests such as the proportion of paired treatment periods in which A outperformed B.[9] When the data are simple and treatment differences are clear, such simple methods work well; graphs are always informative, and plots of the measurements provide good ways to understand the data. But when the number of measurements gets large or when differences are small, graphs will not be sufficient to properly distinguish the treatment effects.

The basic data from an n-of-1 design consist of measurements taken over time while on different treatments. The fundamentals of the statistical analysis can be most easily understood by focusing on the two-treatment design, in which treatments are randomized in blocks of size two, each treatment appearing once in each block. Each treatment period consists of one or more measurement times.

Nonparametric Tests

The earliest n-of-1 trials in medicine used a simple nonparametric test called the sign test. First, one calculates the difference between treatment A and treatment B. If the difference is positive (A is better than B), one counts this as a success. A negative difference counts as a failure. (The choice of which difference is defined to be a success is, of course, arbitrary.) The number of successes, that is, the number of blocks in which A outperforms B, is now compared to the number expected if the treatments were the same: N/2 where N is the number of blocks. Since the number of successes is assumed to follow a binomial distribution, one calculates the probability of the observed result under the null hypothesis that the true success probability is 50 percent. For example, if there were three blocked comparisons and in each A was better than B, the probability would be ½*½*½ = 1/8. This is then a (one-sided) p-value for testing whether A was better than B. This procedure ignores the actual size of the differences and thus ignores potentially important information. Instead, one might use the Wilcoxon signed-rank test on the ranked differences.

While these simple nonparametric tests are easy to use, they ignore important features of the time-series data, particularly their autocorrelation, time trends, and repeated measurements within periods. As a consequence, it is usually worth constructing a proper statistical model that incorporates these features along with an estimation of treatment effect.

Models for Continuous Outcomes

A variety of different models can be constructed when the outcomes are continuous variables, depending on whether they are considered random measurements within each treatment period or vary systematically with time.

First, consider a model in which time may be indexed within treatment periods inside blocks. Notationally, let yijkl represent the outcome measured at time i within treatment period j within block k while on treatment l:

Model 1: $y_{ijkl} = \alpha + \beta_l + \gamma_k + \delta_{j(k)} + \varepsilon_{i(j(k))}$.

Model 1 assumes a fixed treatment effect β_l, random block effects $\gamma_k \sim N(0,\sigma^2_\gamma)$, within-block random period effects $\delta_{j(k)} \sim N(0,\sigma^2_\delta)$, and within-period random errors $\varepsilon_{i(j(k))} \sim N(0,\sigma^2)$, where the notation $N(\mu,\sigma 2)$ indicates a normal distribution with mean μ and variance σ^2. The constant term is used to avoid oversaturation of model terms. Usually, one block is chosen as the reference (e.g., set $\gamma_1 = 0$), and period within-block effects may be expressed so that the difference between the first and second period is assumed the same in each block. This model assumes no time trend and no carryover. The model may be simplified if observations within one treatment period or block are uncorrelated with those in another. In that case, the model becomes a simple two-mean model with random errors:

Model 2: $y_{ijkl} = \beta_l + \varepsilon_{i(j(k))}$

A common scenario for this model is when each treatment period has only one observation (perhaps at the end) to minimize the possibility of carryover.

Modeling Effects Depending on Time

Another class of models pertains to occasions when outcomes vary systematically with time. Causes for such variation include time trends that might describe a disease course or calendar effects that arise from seasonal variation in severity, for instance, in asthma patients whose health is affected by hay fever. Measuring such time effects requires that the study duration and measurement frequency be sufficient to differentiate the trends from noise. It is then easiest to express the model in terms of the measurement y_t taken at time t. If the trend is linear, we have

Model 3: $y_t = \alpha + \beta_t + \gamma X_t + \varepsilon_t$,

in which β is the slope of the time trend, X_t is an indicator for the treatment received at time t, γ is the treatment effect, and ε_t are the residual errors possibly correlated over time. Other time effects can be introduced by modifying the specification for the model, for example, adding nonlinear terms such as quadratic terms to capture possible nonlinear trends. As another example, a seasonal effect could be introduced by adding a dummy variable Z_t taking the value of 1 during the season and 0 outside it. It is important to recognize that the true functional form for time trend is usually unknown; therefore the specification of time effects is usually exploratory.

When each period has a single measurement, the time variable can be replaced by an indicator variable for period. If the effect of treatment is expected to vary with time (e.g., because of higher efficacy during periods of greater disease severity), one can include a time-by-treatment interaction effect into the model.

Autocorrelation

Measurements in a time series typically are not independent, exhibiting some form of autocorrelation that represents the relationship between one measurement and the next in the series. Such autocorrelation arises from time trends or treatment carryover that causes individuals to tend to respond more similarly at times that are closer to each other. Model 3 presents one method of detrending the time series by fitting a model that is linear in time. Such detrending often removes substantial amounts of observed autocorrelation, but some may remain as a consequence of features such as carryover or delayed uptake. Carryover may cause the response to be greater than it should be, if both treatments being compared are active and beneficial. Delayed uptake applies if the full effect of a treatment is not felt at the start of the measurement of the outcome. It will work in the opposite direction, initially depressing the response. The effect of each, however, is to induce correlation between consecutive outcome measurements.

Models that adjust for autocorrelation take two main forms. The first, often called an autoregressive or serial correlation model, expresses the residual error at a given time as a function of the error at one or more previous times, that is, $\varepsilon_t = \delta\varepsilon_{t-1} + u_t$. In this model, δ is the correlation between consecutive errors ε_t and ε_{t-1}. Additional lagged errors of the form ε_{t-k} can be added to the model to represent more complex autocorrelation. The second form, called by some a dynamic model,[10] places the autocorrelation on the outcomes themselves so that the response at time t is a function of the response at time t-1 (and perhaps earlier times). A dynamic form for a model with one fixed treatment effect, for instance, would be $y_t = \delta y_{t-1} + \gamma X_t + \varepsilon_t$. The dynamic model induces a dependence of the current outcome on previous values of the predictors in the model. One can also explicitly introduce this dependence in the form of lagged predictors. It is important to recognize the different interpretation of predictors in a dynamic model resulting from the need to condition on the previous outcome, that is, γ is the treatment effect conditioning on y_{t-1}.

Carryover

Carryover is a special type of autocorrelation common to crossover trials. As stated earlier, it occurs when the time between treatment periods is insufficient for the effect of the previous treatment to end before the next treatment is started. This is common with pharmacological treatments when the drug continues to exert effects in the body after the patient stops taking it. If not controlled for, carryover may lead to bias in the estimated treatment effects, with a tendency to magnify observed treatment effects during transitions from a less effective (but still effective) treatment to a more effective treatment, and conversely to shrink effects during transitions from a more effective to a less effective treatment.

Both design and analytic approaches can address carryover. Designing washout periods long enough for the prior treatment's effect to disappear by the beginning of the next treatment period eliminates any potential correlation across periods. An analytic approach downweights, disregards, or simply does not collect outcomes at the beginning of a treatment period, thus creating an analytic washout period.[11] This analytic approach is also helpful when treatments take time to reach their full effect and one desires to account for the reduced effect at the beginning of the period.

Zucker[12] used an extreme version of this approach in a series of n-of-1 trials for patients with fibromyalgia tested on amitryptoline or amitryptoline plus fluoxetene. Treatment periods were 6 weeks long, and the primary outcome was the score on the patient-reported Fibromyalgia Impact Questionnaire. Only the report from the end of each treatment period was analyzed. While this almost certainly eliminated carryover, and in fact autocorrelation, it did have the drawback of giving only one measurement per treatment period. In some studies, however, these choices may be unavailable if each treatment period is short or treatment half-life is very long.

Various approaches to estimating carryover have been proposed. As Senn[13] points out, all rely on restrictive modeling assumptions and are inferior to designing a proper washout (which also may rely on assumptions about pharmacologic or similar properties of the treatments). The discussion above points to autocorrelation models as one method to handle carryover, although they assume correlations over time unrelated to when treatment is changed or introduced. In principle, one could design an autocorrelation structure that varied with time since introduction of treatment. But this would need to assume characteristics of the nature of the carryover that might not be well supported.

A simple check for carryover when the analyst has a sufficient number of observations taken over time within each treatment period is to compare results using all measurements to results after discarding those at the beginning of the period that might be affected by carryover. The model with more measurements should return more precise estimates but at the risk of some bias from the carryover. If the estimates are similar, carryover is not likely to be an issue.

Another form of carryover that one might be able to examine is the effect of treatment sequence when the response is different depending on the order of the treatments given. Treatment A may have a bigger effect if given after treatment B. This might manifest itself through responses that are higher for treatment A when it follows B than when it follows another period of A. One can examine a sequence effect by adding a variable that codes for sequence, for example, a dummy variable that equals 1 in periods where A follows B and 0 otherwise. Of course, if treatment effects are wearing off, it would not be appropriate to code every measurement in the A period with the sequence effect.

Discrete Outcomes

In each of the models presented, we have assumed a continuous outcome with normally distributed measurement error. Many outcomes that might be used in n-of-1 trials, however, may use categorical scales, event counts, or binary indicators of health status. For example, Guyatt[14] and Larson[15] both used Likert scales with ratings from 1–7 to measure patient outcomes. Models for such outcomes require different formulations that do not rely on the assumption of normality.

Generally, one needs to formulate such models as generalized linear models.[16] Binary outcomes use logistic regression; count outcomes use Poisson regression; and categorical outcomes use categorical logistic regression. The generalized linear model has the same form as the linear model on the right-hand sides of the models above, but expresses the left-hand side in terms of a (link) function of the mean of the probability distribution for the outcomes. For example, with a binary outcome, events occur according to Bernoulli distribution, and the mean of that distribution is the probability of an event. The link function used in logistic regression is the logit function ($\text{logit}(p) = \log_e(p/(1-p))$). In Poisson regression, the link function is log. For categorical regression, various link functions can be used depending on how one wants to model the data. A common link function for an ordered outcome such as a preference scale is the cumulative logit.[17]

Although the generalized linear models use different estimation algorithms and take different functional forms, model construction does not differ conceptually in any fundamental way from the normal linear models, so we will say no more about them here, but refer the interested reader to the many textbooks that treat them.[16,17]

Estimation

The simplest approach to estimating the treatment effect is based on the model that ignores any potential effects of time, autocorrelation, or carryover and simply compares the average response when the patient is on each treatment. If the design is blocked, one can take the difference between outcomes within each block and then simply average the differences, computing the appropriate standard error. This corresponds to a paired t-test. If no blocking is used, the analysis is an unpaired t-test.[13]

In general, one can use likelihood methods that incorporate the necessary correlation structures and interaction terms to fit the models. Likelihood-based methods typically rely on large samples to validate their assumptions of normal distributions of the resulting model estimates. Because the amount of data collected on any single outcome in an n-of-1 study is small, such assumptions may not be appropriate.

Bayesian inference combines this likelihood with prior information to form a posterior distribution of the likelihood that a model parameter takes a given value. The prior information is expressed through a probability distribution describing our degree of belief about model parameters before observing the data. Bayesian inference is natural for clinicians making decisions such as a differential diagnosis, because it expresses the way that they combine new information (such as a diagnostic test result) to update their previous beliefs.[18] In an n-of-1 trial, the prior may be based on a population average effect or may be individualized to reflect patient-specific characteristics. The use of prior information also permits the analysis to incorporate patient preferences and beliefs.

Specification of a complete prior distribution for all model parameters can be difficult, particularly for those, such as correlations or variance components, about which not much may be known. One common simplification assumes that very little is known about some or all of the parameters and uses prior distributions that do not favor any values over others. Probabilistically, this corresponds to a uniform (flat) distribution. Such priors are referred to as noninformative. Conversely, knowledge of certain parameters such as the expected treatment effect may be available, and so informative priors may be chosen. For example, for a pain scale outcome the average pain reduction that one can expect over a 2-week course of therapy may be approximately known in the population, or one may be able to bound the maximum amount. It is also possible to construct an approximate prior distribution by eliciting its key parameters, such as its mean and standard deviation, or its percentiles.[19]

The posterior distribution, formed by calculating the conditional probability distribution of each parameter given the observed data and the specified prior distribution, is essentially a weighted average of the observed treatment effect mean and the hypothesized prior mean. The weights are supplied by the relative information about the two expressed through the precision with which each is known. One can use the posterior distribution to make statements about the probability that the parameters take on different values. For instance, one might conclude that the chance that treatment A reduces pain more than treatment B as measured on a specific pain scale is 75 percent, or one might say that there is 50-percent chance that the reduction is at least 10 points on the scale. Statements like this can be made for each outcome, allowing the patient and clinician to weigh them and determine which treatment is working better. Bayesian inference leads to statements about the probability of different hypotheses given the data observed; non-Bayesian, or frequentist, inference leads to statements about the probability of the data given the null hypothesis.

Local Knowledge and Statistical Methods

The personalized nature of n-of-1 trials indicates that the primary use for the knowledge produced in each individual trial is to inform clinical decisionmaking for the specific patient, that is, the knowledge produced is used locally or internally within the patient-clinician team that produced it. This paradigm is crucially different from the situation in standard parallel group randomized controlled trials (RCTs), in which the primary use of the knowledge produced in an RCT is to inform clinical decisionmaking for future patients, rather than for the patients participating in the RCT. In fact, for double-blinded RCTs, the patients and their clinicians do not know what treatment the patient actually received until the RCT is unblinded. Given this fundamental difference between the two paradigms, the appropriate statistical method also differs. While significance testing is the usual statistical method for the standard parallel group RCTs, the same method might be less pertinent for n-of-1 trials. Instead, one provides the decisionmaker with all the information required in a format that facilitates decisionmaking.

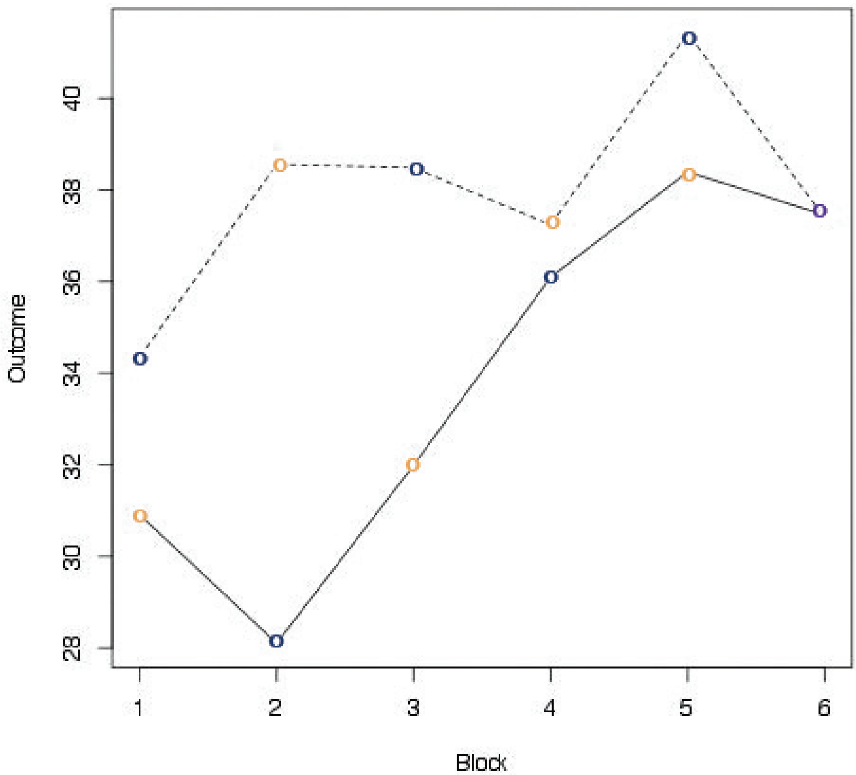
Presentation of Results

In order to make a correct decision, it is important that the patient and clinician not only have the right information, but that it be presented to them in a format that is easy to understand. The results of a trial are complex, and data are collected on multiple outcomes at many times under different treatment conditions. Many of the models we have discussed describe complicated phenomena such as autocorrelation that may confound facile interpretation of the data. Further complications might be present in skewed and/or heteroskedastic data (such as lognormal data and Poisson-distributed count data) that might indicate transformation to a different scale for graphic presentation and statistical modeling. Good graphics can help explain the data and the results to all parties involved.

Results should always be accompanied by the simplest possible graph, plotting each outcome over time separately in the treatment and control groups. A variety of different approaches are possible. One could overlay or stack two line plots, matching by block pairs. This reveals within-block differences as well as time trends and potential autocorrelation. One could add the sequence order by separately coloring within each block the first sequence in one color and the second in another (as in Figure 4–1). Such displays of raw data provide important information on the relationship of outcomes to treatment. They may also be shown in a blinded fashion (without identification of treatment group) to the patient during the trial as a form of feedback to motivate adherence.

Kratchowill et al.[20] describe a process for using figures to evaluate the success of the intervention. After establishing a predictable baseline pattern of data, one examines the data within each phase to assess the performance and potentially to extrapolate to the next phase. Assessment involves: (1) level, the mean for the data in a phase; (2) trend, the slope of the line of best fit in a phase; (3) variability of the data around this line; (4) immediacy of the effect, the change in level between the end of one phase and the next; (5) overlap, the degree to which the data at the end of one period resemble those at the beginning of the next; and (6) consistency of data patterns in similar phases. More consistency, separation between phases, and strong patterns suggest a real effect. Once each phase is assessed, results from successive phases are compared to determine if the intervention had an effect by changing the outcome from phase to phase. Finally, one integrates the information across phases to see if the effects are consistent. A similar scheme is given in Janovsky.[21]

Figure 4–1. Data from simulated n-of-1 trial



Note: Two line plots (solid and dotted) show outcomes for 2 treatments measured within each of 6 blocks. Patients receive each treatment in each block, with the point labeled in orange taken first.

Determining treatment differences directly from such figures may, however, be camouflaged by other features of the data such as autocorrelation and time trends. Figure 4–1 shows simulated data showing that treatment B (dotted line) typically produces better outcomes than treatment A (solid line). Responses appear to be increasing with time on treatment A, but not B, suggesting a potential treatment-by-block interaction. Because only one measurement is recorded on each treatment period, we cannot distinguish time effects from effects by block. The overall effect of the picture is that B may be better than A, but that this efficacy wears off over time. In fact, the data are simulated with a fixed treatment difference and with a trend over time, but no treatment-by-block interaction, which occurs by chance. The right answer (discernible through use of appropriate statistical analysis) is that B is better than A and that all patient responses are increasing with time. Therefore, the plot is somewhat misleading and may prompt the wrong decision. As a general rule, unless treatment effects are large or specific, plots will provide necessary but not sufficient information to make appropriate decisions. It is therefore important to supplement graphs with appropriate statistical analysis and present the information in the clearest way possible.

One should use the statistic provided by the modeling process that relates directly to the measured treatment difference. In the Bayesian framework, this is the posterior probability; in the non-Bayesian, or frequentist, framework, this is typically a p-value. We recommend the Bayesian approach because it provides more value to the patient. The p-value describes the likelihood of obtaining the actual data (or more extreme data) under a specific null hypothesis. For example, a p-value of 0.05 for a test of the null hypothesis of no difference in treatments means that if the two treatments had the same effect, one would have observed the difference found (or a more extreme difference) 1 time in 20 times under repeated sampling. Putting aside the irrelevancy of the repeated sampling assumption (since the experiment will not be repeated), one is left with the observation that it is unlikely that the treatments have the same effect, but one does not know the likelihood of any other effect.

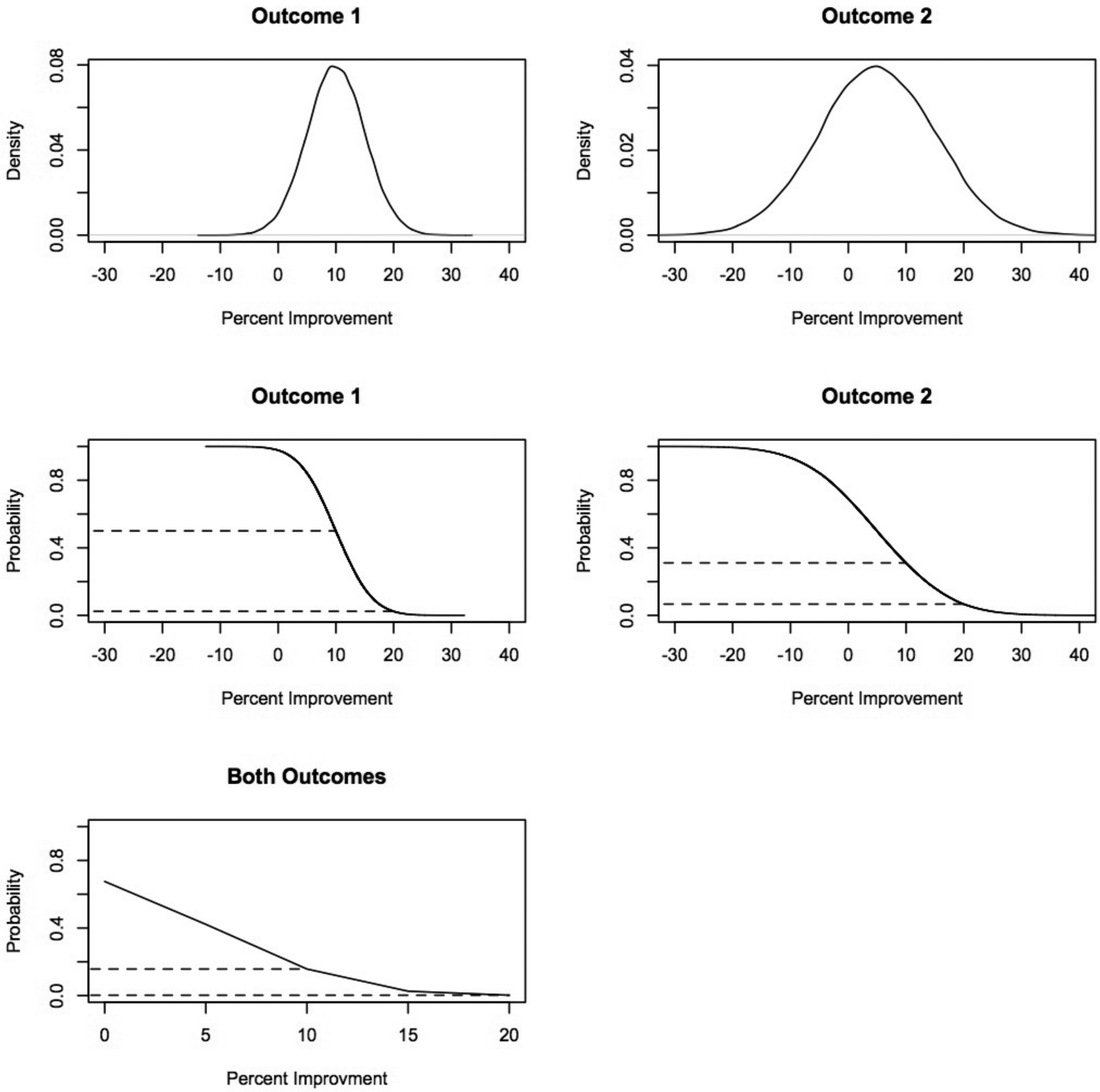
Contrast this with the Bayesian interpretation, which gives the full posterior probability distribution of the treatment effect under the model chosen. From this posterior distribution, one can make probabilistic statements about the likelihood of any size of treatment effect, for example, the likelihood that the treatment effect is at least 10, or between 5 and 15. In essence, this approach focuses on estimation of the magnitude of the effect, rather than on hypothesis testing. As a result of the focus on estimation instead of hypothesis testing, power analysis is of less concern. Zucker et al.,[8] quoted in Duan et al.,[22] show that for a study with M patients and N paired-time periods, study precision is $M/(\tau^2 + 2\sigma^2/N)$, which provides a way to calculate the tradeoff in sample size between the patients and time periods.

This focus can be particularly informative when multiple outcomes are of interest to the patient, and one wants to balance different objectives. As an alternative to the use of the composite scale discussed previously, one could formulate a joint posterior distribution to make probabilistic statements about the joint probabilities attached to combinations of the outcomes, if one were prepared to make some assumptions about their relationships. As an example, assume that the users (the patient and clinician) specified a performance target for the new treatment, A, to improve pain by at least 10 percent and increase sleep by at least 1 hour per night compared to the current treatment, B. In the simple (and perhaps unrealistic) case that the outcomes are independent, the probability for the joint outcome is the product of the probabilities of each separate outcome. So, if the probability that A improved pain by 10 percent was 0.3 and the probability that A increased sleep by 1 hour was 0.2, then the probability that both would happen would be 0.06.

Such probabilities can be expressed by a distribution function of the likelihood of each gain or by a cumulative distribution. As an example, assume that the posterior distribution of treatment benefit on A compared to B for outcome 1 expressed as a difference in percent change from baseline was normally distributed with mean 10 percent and standard deviation 5 percent. Therefore, there is roughly a 97.5-percent probability that A has bigger benefit than B, since 0 change is about two standard deviations below the mean. Likewise, assume the benefit for the second outcome is smaller but more uncertain, normally distributed with mean 5 and standard deviation 10. Figure 4–2 (top row) plots the resulting posterior probability distributions of treatment effect for each outcome together. One might also be interested in their cumulative distributions, or more likely, the probability of observing an improvement at least as big as a certain size. These graphs appear in the middle row of the figure. Using the dotted lines on the graph, we can see that the probability of at least a 10-percent improvement is slightly higher with outcome 1 than with outcome 2 since its mean is higher, but that the situation reverses for the probability of at least a 20-percent improvement because of the greater uncertainty associated with outcome 2. The bottom row of the figure gives the probability that both outcomes are improved by a given amount. This probability is smaller than for either outcome alone and (for this example) is roughly the product of the two individual probabilities, because the two outcomes were simulated independently. In practice, these joint probabilities may be quite similar to or quite different from their components, depending upon the correlation between the outcomes.

While plots like those in Figure 4–2 display the entire distribution of effect sizes along with our uncertainty in estimating them, some may prefer a simpler display with less total information, but perhaps in a format that is easier to understand. The distributions in the top row of the figure may be collapsed into a median and a central interval displaying the values most likely to occur with a given amount of probability, often 95 percent. One may also choose one or more increments of

Figure 4–2. Percent improvement and probability of improvement for two outcomes



Note: Top row: Posterior distributions in percent improvement (treatment effect) for 2 outcomes; Middle row: Probability that outcome improves by at least amount on horizontal axis for each outcome; Bottom row: Probability that both outcomes improve by at least amount on horizontal axis.

improvement for which to display probabilities. Figure 4–3 displays the median and 95-percent central interval (from the 2.5 to the 97.5 percentile) for the treatment effect for each outcome. The probabilities associated with improvement of at least 0, 5, 10, 15, and 20 percent for each outcome and both outcomes together can be displayed as in Table 4–1. The users should be able to specify the exact outcome levels for which they want probabilities computed. These may correspond, for instance, to clinically relevant values as determined by the patient and clinician in collaboration.

Figure 4–3. Posterior median and 95-percent central posterior density interval for two outcomes

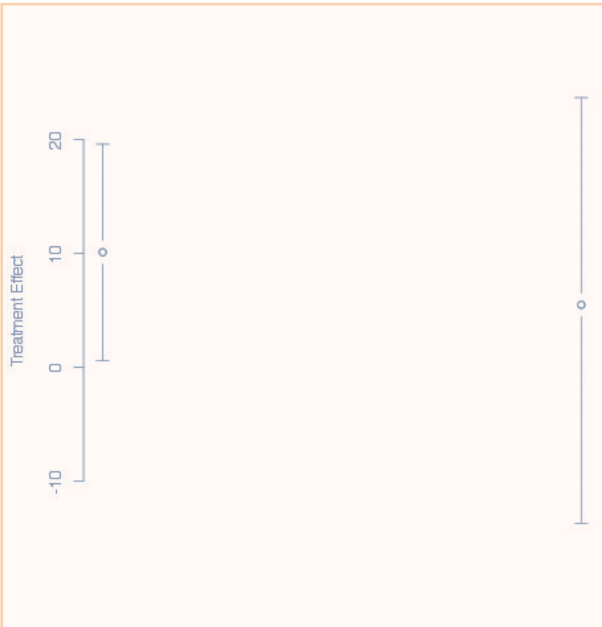


Table 4–1. Probability that given outcome or two outcomes together have a treatment effect greater than a given amount

Exceedance Probability	Outcome		
	1	2	1 and 2
Probability > 0	0.97	0.69	0.67
Probability > 5	0.86	0.50	0.43
Probability > 10	0.51	0.31	0.17
Probability > 15	0.17	0.16	0.02
Probability > 20	0.02	0.07	0.00

Some users may prefer to consider results as odds, rather than probabilities. Others may prefer metrics other than treatment effects. A flexible environment in which the user can request results in the way that is most comfortable and personally informative is a desired feature of any n-of-1 analytic module.

Combining N-of-1 Studies

Although n-of-1 studies are designed for single patients working with a single clinician to make a single treatment decision, many n-of-1 studies may be similar enough to inform others. Furthermore, the small number of crossovers used in many n-of-1 studies may increase the need to combine the index patient's data with data from other patients who participated in similar n-of-1 trials to increase the statistical precision available for making individual treatment decisions.

Such similarity may arise from the same clinician testing the same treatments with patients having the same condition; similar patients testing the same treatments with different clinicians; clinicians within the same clinic practicing in similar ways; or examining a common set of treatments in different combinations. In each case, we may think of the set of n-of-1 studies as forming a meta-analysis and attempt to combine them using techniques from meta-analysis such as multilevel random-effects models, regression, and networks. As an added bonus, combining the results can help estimate the average treatment effect in the population as well as the individual treatment effects for single patients. We give a brief introduction here, but refer the interested reader to related treatments in Zucker[23] and Duan, Kravitz, and Schmid.[22]

To extend the previous models to multiple patients with n-of-1 studies, we assume that the same outcome measure, y, is used across patients to be combined, and consider

Model 1a:, $y_{mijkl} = \alpha_m + \beta_l + \gamma_k + \delta_{j(k)} + \varepsilon_{i(j(k(m)))}$

where m indexes the patient, $\alpha_m \sim N(0,\sigma^2_\alpha)$ is the random effect for the patient, and the error term indicates the variability within observations taken within a treatment period within a block within a patient. The time-trend model,

Model 3a: $y_{it} = \alpha_i + \beta_t + \gamma X_t + \varepsilon_{it}$,

changes only by having a random intercept $\alpha_i \sim N(0,\sigma^2_\alpha)$ for the i-th patient. These models may be easily extended to encompass interactions between patients and other factors that would indicate variation across patients. In particular, patient characteristics may explain some of the between-patient variance σ^2_α.

If we assume all within-block measurements are exchangeable, that is, all block-specific treatment effect estimates are similar and can be considered replicates of each other, we can combine results across patients quite simply. First, estimate the treatment effect for patient m within block b as the difference in the outcomes between treatment 1 and treatment 0, $D_{mb} = \sum_{i,j}(Y_{mijb1} - Y_{mijb0})$. The block-specific treatment effect estimates can then be aggregated across blocks to form the individual treatment effect (ITE) estimate $\bar{D}_m = \sum_{b=1}^{B_m} D_{mb}/B_m$. It is possible to extend this approach to a regression estimate under the broader assumption that allows observed differences across blocks, such as a period effect. The observed ITEs $\bar{D}_m$ are unbiased for the true ITE Φ_i, so that $\bar{D}_m \sim N(\Phi_i, s_i^2)$. The within-patient variance s_i^2 is assumed to be known and allowed to be specific to each patient (as in a meta-analysis treating each patient as a study). This permits capture of variation in design or implementation of the studies, such as the variation in the number of blocks across patients. For instance, one could assume that $s_i^2 = \sigma^2/B_i$ equals the common within-block variance σ^2 scaled by the number of blocks. If the full model 1 is used, then s_i^2 is estimated from the within-block measurements.

The true ITEs are assumed to be drawn from a random-effects distribution, $\Phi_1 \sim N(\Phi_0, \tau^2)$, where Φ_0 denotes the overall mean treatment effect for the population, and τ^2 denotes between-patient variance in the individual mean treatment effects. Prior distributions are placed on the parameters Φ_0, τ^2, and σ^2 to represent what is known about these parameters prior to the study. The overall mean treatment effect Φ_0 and the individual mean treatment effects Φ_i's are estimated using the posterior distribution for each parameter.

The posterior distribution of the patient's ITE, Φ_i, provides an opportunity to obtain a more informative estimate of the ITE than is available in a single n-of-1 trial because of the opportunity to borrow strength from the population mean Φ_0. Recall that the posterior mean is an average of the sample mean and the prior mean. In this situation, the prior mean Φ_0 is the external information coming from other patients and $\bar{D}_m$ is the information coming from the patient. If the patient is like the others, the posterior mean will be close to the average.

The relationship between individual treatment effect, Φ_i, and overall treatment effect, Φ_0, depends on the balance between the between-patient variance, τ^2, and the within-patient variance, s_i^2.[24] When between-patient variance is small compared to within-patient variance (i.e., little or no heterogeneity of treatment effects), the patient-specific mean treatment effects, Φ_i, are very similar and close to the posterior mean effect, Φ_0. Alternatively, if between-patient variance is large compared to within-patient variance (i.e., strong heterogeneity of treatment effects), the Φ_i would be estimated to be close to the patient-specific treatment effect estimate, $\bar{D}_m$, with little or no "borrowing from strength." In a sense, because the "strength" (population information) to be borrowed does not provide strong statistical information, within-patient information dominates between-patient information.

The model for multiple patients may be extended by considering the model as comprising two parts, within-patient and between-patient. The models for the single n-of-1 trial describe the within-patient parts. The between-patient parts describe factors that vary among patients, as in any statistical model with patient units. These include patient characteristics such as comorbidity, demographics, and socioeconomic status. They may also include study and health care structure such as the nesting of patients within providers and providers within organizations. Each level in the nested structure is represented by a random effect, in addition to the patient-level random effect Φ_i. For example, the model that accommodates a nested structure with patients nested within practices will have a random effect for practices in addition to a random effect for patients: $\Phi_{pi} \sim N(\theta p, \tau_p^2)$ with $\theta_p \sim N(\theta_0, \omega^2)$ where Φ_{pi} denotes the individual mean treatment effect for the i^{th} patient in the p^{th} practice, θ_p denotes the mean treatment effect among patients in the p^{th} practice, τ_p^2 denotes the within-practice variance among patients in the p^{th} practice, θ_0 denotes the overall mean treatment effect across practices, and ω^2 denotes the variance across practices. Again, covariates at the practice level can also be incorporated into the model to evaluate the heterogeneity of treatment effects (HTE) associated with these covariates.

In addition to improving estimates of a patient's ITE through borrowing strength from other studies, one also obtains an estimate of the overall treatment effect across patients either as a single mean or as a regression. These population effects can be used to inform treatment decisions for similar patients who did not participate in n-of-1 trials.

Finally, when n-of-1 trials with different treatment comparisons are combined across patients, it is possible to consider a network meta-analysis of the n-of-1 trials. Models for network meta-analysis[25,26] incorporate all the pairwise comparisons into a single model for simultaneous estimation. Under assumptions of consistency[27] and similarity,[25,28] direct comparisons of treatments A and B, A and C, and B and C may be combined so as to incorporate both their direct estimates and indirect estimates. (AC is estimated indirectly through the sum of AB and BC.) Such models make optimal use of all the treatment data, leading to more precision in effect estimates as well as the ability to rank treatments. These models hold even when studies do not compare all treatments, but only a subset. For example, a study comparing A and B may be combined with one comparing B and C to get an indirect estimate of A and C. Studies with more than two arms not only fit into the model structure, but actually provide additional information, because their direct and indirect estimates obtained from the same study must be consistent.

Automation of Statistical Modeling and Analysis Procedures

The implementation of the statistical modeling and analysis procedures, including procedures for a single n-of-1 trial and procedures for combining n-of-1 trials across patients, needs to be facilitated by building these procedures into the informatics system to allow for automation of these procedures, in conjunction with periodic review and calibration of the procedures for continuous quality improvement. Such automation is particularly important because most clinician/patient pairs will have neither the time nor the expertise to evaluate the statistical models, and rarely will a statistician be available to do the analysis in real time. Instead, it will be necessary to have the statistical modeling, including model selection, model checking, and model interpretation, built into the informatics system for presentation when a treatment decision is to be made.

Conclusion

N-of-1 data offer rich possibilities for statistical analysis of individual treatment effects. The more data that are available both within and across patients, the more flexibility models have. This richness does come at the price of the need for careful model exploration and checking. Many errors can be avoided with good study design that respects standard experimental principles and minimizes the risk of complexity caused by autocorrelation, as by including washout periods to minimize carryover. Such design and modeling expertise is probably not within the realm of the average clinician and patient undertaking an n-of-1 study. Thus, it is crucial that standard protocols and analyses be available, especially in an automated and computerized format that promotes ease of use and robust designs and models.

Checklist

Guidance	Key Considerations	Check
Treatment assignment needs to be balanced across treatment conditions, using either randomization or counterbalancing, along with blocking	• Design needs to eliminate or mitigate potential confounding effects such as a time trend. • Pros and cons of randomization versus counterbalancing need to be considered carefully and selected appropriately. Counterbalancing is more effective if there is good information on critical confounding effect, for example, linear time trend. Randomization is more robust against unknown sources of confounding. • Blocking helps mitigate potential confounding with time trend, especially when early termination occurs.	❒
Blind treatment assignment when feasible	• Blinding of patients and clinicians, to the extent feasible, is particularly important for n-of-1 trials, especially with self-reported outcomes, when it is deemed necessary to eliminate or mitigate nonspecific effects ancillary to treatment. • Some nonspecific effects might continue beyond the end of trial within the individual patient, and therefore should be considered part of the treatment effect instead of a source of confounding.	❒
Invoke appropriate measures to deal with potential bias due to carryover and slow onset effects	• A washout period is commonly used to mitigate carryover effect. • Adverse interaction among treatments being compared indicates the need for a washout period. • Absence of active treatment during a washout period might pose an ethical dilemma and diminish user acceptance for active control trials. • Washout does not deal with slow onset of new treatment and might actually extend the duration of transition between treatment conditions. • Analytic methods can be useful for dealing with carryover and slow onset effects when repeated assessments are available within treatment periods.	❒
Perform multiple assessments within treatment periods	• Repeated assessments within treatment periods can enhance statistical information (precision of estimated treatment effect) and facilitate statistical approaches to address carryover and slow onset effects. • The costs and respondent burden need to be taken into consideration in decisions regarding frequency of assessments.	❒

Checklist (continued)

Guidance	Key Considerations	Check
Consider adaptive trial designs and sequential stopping rules	• Adaptive trial designs and sequential stopping rules can help improve trial efficiency and reduce patients' exposure to the inferior treatment condition.	❐
Use appropriate statistical method to analyze outcome data, taking into consideration important features of time-series data, including autocorrelation, time trend, and repeated measures within treatment periods	• Mixed-effect models, autoregressive models, and dynamic models can be used to analyze time-series data from n-of-1 trials. • Nonparametric tests are easy to use but might not fully capture time-series features. • Significance testing is less pertinent for n-of-1 trials than the provision of the information needed for the users to make decisions for future treatments.	❐
Use appropriate methods to handle multiple outcomes	• Separate analyses and reporting of trial findings for multiple outcomes can accommodate the patient-centered nature of n-of-1 trials. • Explicit prespecification of weights across outcomes is preferable to post hoc weighting. • A composite index or scale can effectively synthesize information across related outcomes and reduce the burden on users to digest trial results across multiple outcomes.	❐
Present results of statistical analysis in an informative and user-friendly manner	• Customize format of presentation to accommodate needs and preferences for individual users. • Graphical presentation of trial results is easy to comprehend but might be complicated by autocorrelation, time trend, etc. • Posterior probabilities or odds based on a Bayesian framework are more interpretable for users than p-values based on a frequentist framework.	❐
Borrow from strength	• Bayesian methods can be used to combine data across individuals participating in similar n-of-1 trials, to provide more precise estimates for individual treatment effects, and also to provide estimates for average treatment effects in the population to inform treatment decisions for patients not in the trials. • Network meta-analysis can be used to incorporate information from patients whose trials are related to but not identical in design to the treatment conditions compared.	❐

References

1. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: RandMcNally; 1963.
2. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Belmont, CA: Wadsworth; 2002.
3. Ross M, Olson JM. An expectancy-attribution model of the effects of placebos. Psychology Review. 1981;88(5):408-437.
4. Rutherford BR, Marcus SM, Wang P, et al. A randomized, prospective pilot study of patient expectancy and antidepressent outcome. Psychological Medicine. 2013;43(5):975-982.
5. Berry SM, Carlin BP, Lee JJ, et al. Bayesian adaptive methods for clinical trials. NY: CRC Press; 2010.
6. Zelen M. Play the winner rule and the controlled clinical trial. Journal of the American Statistical Association. 1969;64(325):131-146.
7. Wei LJ, Durham S. The randomized play-the-winner rule in medical trials. Journal of the American Statistical Association. 1978;73(364):840-843.
8. Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. Journal of Clinical Epidemiology. Dec 2010;63(12):1312-1323.
9. Gabler NB, Duan N, Vohra S, et al. N-of-1 trials in the medical literature: a systematic review. Medical Care. Aug 2011;49(8):761-768.
10. Schmid CH. Marginal and dynamic regression models for longitudinal data. Statistics in Medicine. 2001;20(21):3295-3311.
11. Hogben L, Sim M. The self-controlled and self-recorded clinical trial for low-grade morbidity. British Journal of Preventive and Social Medicine. 1953 7(4):163-179. Reprinted in International Journal of Epidemiology 2011; 40(6):1438-1454.
12. Zucker DR, Ruthazer R, Schmid CH, et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. Journal of Rheumatology. 2006;33(10):2069-2077.
13. Senn S. Cross-over trials in clinical research. 2nd ed. Hoboken, NJ: Wiley; 2002.
14. Guyatt GH, Keller JL, Jaeschke R, et al. The n-of-1 randomized controlled trial: clinical usefulness. Our 3-year experience. Annals of Internal Medicine. 1990;112(4):293-299.
15. Larson EB, Ellsworth AJ, Oas J. Randomized clinical-trials in single patients during a 2-year period. Journal of the American Medical Association. 1993;270(22):2708-2712.
16. McCullough P, Nelder J. Generalized linear models. 2nd ed. NY: Chapman and Hall; 1989.
17. Agresti A. Categorical data analysis. 2nd ed. Hoboken, NJ: Wiley; 2002.
18. Gill CJ, Savin L, Schmid CH. Why clinicians are natural Bayesians. Bristish Medical Journal. 2005;330(7499):1080-1083.
19. Chaloner K, Church T, Louis TA, et al. Graphical elicitation of a prior distribution for a clinical-trial. Journal of the Royal Statisticial Society. Series D (The Statistician). 1993;42(4):341-353.
20. Kratochwill TR, Hitchcock J, Horner RH, et al. Single-case designs technical documentation. 2010;2013.
21. Janosky JE, Leininger SL, Hoerger MP, et al. Single subject designs in biomedicine. Springer 2009
22. Duan N, Kravitz R, Schmid C. Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. Journal of Clinical Epidemiology. 2013;66(Suppl 1): S21-S28.
23. Zucker DR, Schmid CH, McIntosh MW, et al. Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. Journal of Clinical Epidemiology. 1997;50(4):401-410.
24. Schmid CH, Brown EN. Bayesian hierarchical models. Methods Enzymol. 2000;321:305-330.
25. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Research Synthesis Methods. 2012;3(2):80-97.
26. Higgins J, Jackson D, Barrett J, et al. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Research Synthesis Methods. 2012;3(2):98-110.

27. Lu GB, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. Journal of the American Statistical Association. 2006;101(474):447-459.

28. Jansen JP, Schmid CH, Salanti G. Directed acyclic graphs can help understand bias in indirect and mixed treatment comparisons. Journal of Clinical Epidemiology. 2012;65(7):798-807.

Chapter 5. Information Technology Infrastructure for N-of-1 Trials

Ian Eslick, Ph.D.
MIT Media Laboratory

Ida Sim, M.D., Ph.D.
University of California San Francisco

The DEcIDE Methods Center N-of-1 Guidance Panel*

Introduction

N-of-1 trials have not been adopted for broad use despite early successes and the promise of improved care and reduced costs.[1] Barriers to adoption include education (addressed in Chapter 6), operational complexity,[2] and costs. For example, in one system, the time spent per trial was approximately 16.75 hours. Half that time was spent in setup and another third for trial execution.[3] Some of these time costs were attributable to carrying out the n-of-1 process, involving direct patient education and discussion. The remaining time was spent on activities related to trial design, transmission of design to the pharmacist, analysis, preparation, and presentation of results. A detailed discussion of n-of-1 costs can be found in Chapter 3.

This chapter describes how a modern information technology (IT) infrastructure and design approach can reduce the costs, burden to end-users (patients and their clinicians), and complexity of administering and running n-of-1 trials at scale by automating trial workflow. IT infrastructure also offers new opportunities such as the ability to pull data from electronic medical records (EMRs), integrate with emerging consumer health devices, interact with patients or collect data via mobile platforms, and embed statistical analysis and visualization directly into Web-enabled platforms.

Prior work has addressed the procedures necessary to run individual trials using pen-and-paper techniques or simple electronic tools such as a spreadsheet,[3] but to date attempts to build IT platforms for n-of-1 trials have not been commercially successful (see Chapter 3). Unfortunately, existing clinical trial management systems are inadequate for managing n-of-1 trials and are difficult to extend for this purpose. This chapter discusses features of an IT-based trial platform that will enable efficient scaling beyond individual provider use. It is written to provide clinicians, health services researchers, and IT specialists a shared framework for discussing the use of IT to support n-of-1 trials. It defines relevant terms, identifies key requirements of n-of-1 trial systems, introduces tradeoffs to be considered, and warns of common pitfalls to avoid. While all of the proposed features should be considered during a project's design phase, only a subset are likely to be implemented in any one platform due to the complexity inherent in modern health care IT systems.

A research IT system, called MyIBD, is presented as an early example of such an integrated system. MyIBD is a prototype of a Personalized Learning System that uses longitudinal data collected from the context of everyday life to inform the management of chronic disease. Support of individual n-of-1 trials was one of the design goals for the platform. A pilot study by four health services researchers engaged a small pool of providers from three hospitals toward a target of 20 concurrent patients to identify issues involved in performing and scaling IT-supported n-of-1 trials as an intrinsic part of the care of chronic diseases. A future revision of the platform, informed by the pilot, is intended for deployment in the 50 or

*Please see author list in the back of this User's Guide for a full listing of panel members and affiliations.

more centers of the ImproveCareNow[4] pediatric inflammatory bowel disease (IBD) quality improvement network.

This chapter should be read as providing an array of practical options for organizations looking to develop an n-of-1 trial platform. As commercial and open-source IT platforms become available, this document will help organizations evaluate these offerings.

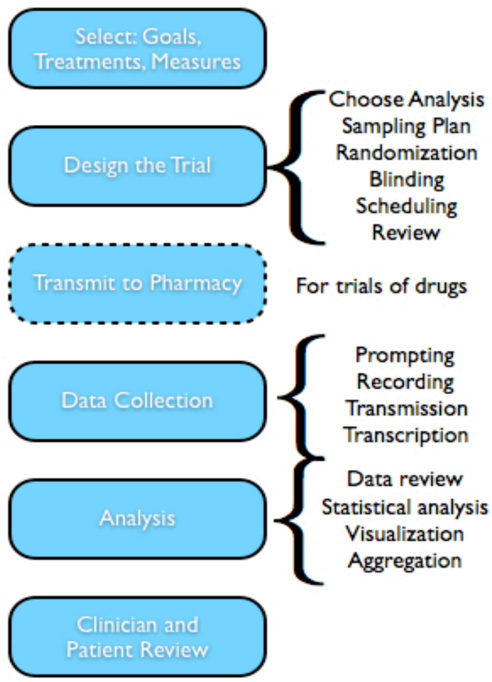
What Does an N-of-1 Trial Platform Do?

A general trials administration platform facilitates all phases of design, execution, and analysis activity as illustrated by Figure 5–1. Compared with traditional clinical trials, n-of-1 trials allow for greater user (patient and provider) participation in the selection and design phases through a dialog between a health professional and a patient and/or family member. The design of an n-of-1 trial may be specified by the care provider independently, through a shared decisionmaking exercise between the patient and the provider, or by a patient-driven process, to determine the treatments to compare, the outcomes to track, the format and content of the final report to be presented at the end of the trial, etc. In all cases, IT can support these steps by providing ready access to standard libraries of characterized treatments, outcome measures, and statistical tools. IT can help clinicians and patients jointly explore tradeoffs that affect the time, strength, and overall burden of the trial.

The end result of the design phase is a schedule of treatment periods with specified treatments and measurements necessary to execute the trial. An electronic platform can facilitate the trial's execution through data collection, treatment reminders, and pharmacy interaction where necessary. Some trials may involve action plans that accommodate real-world challenges such as acute comorbidity, changes in routine, or periods of nonadherence. The platform should support a variety of adjustments to trials and track these adjustments for use in subsequent analyses.

When the trial is complete, the platform should execute statistical analyses as specified in the protocol, which may include adaptive model selection; and the results should be displayed for joint clinician-patient interpretation (see Chapter 4). Because nearly all n-of-1 trials are conducted to inform a treatment decision in the ongoing care of a patient, visualizations and reports must be understandable to patients and their clinicians. Shared decisionmaking aids are an important part of facilitating the use and interpretation of n-of-1 trials in clinical care.

Figure 5–1. Platform processes



At scale, many individual trials are likely to be variations on a common theme. A library of experimental designs should also be supported by the platform. The track records for prior trials using a specific design can inform future users of problems with the protocol, the rate of successful trial completion, the ability to reach a definitive conclusion, and user feedback on their experience. Providers using the same protocol can exchange notes on or outside the platform. An automated scientific review process could be added to this IT infrastructure to ensure the soundness of the protocol and analysis plan for each trial.

Table 5–1. Possible user roles in an n-of-1 trial platform

Patient and/or Caregiver	Access own data. Codesign n-of-1 trial. Enter data via Web, Short Message Service (SMS), mobile app, device, third-party service. View and interpret results.
Clinician	Recruit and manage a sample of patients. Codesign n-of-1 trials. Monitor trial and data collection progress, and intervene as needed. View and interpret results.
Administrator	Provide institutional oversight, user account creation, and management.
Pharmacist	Receive instructions for specific patients/trials, including randomization schedule and blinding requirements. May interact through the trial system or by fax, secure email.
Statistician/ Researcher	Review trial design and collected data for validity and/or aggregate analysis. May download identified or de-identified data for offline analysis.
Systems Administrator	Support operation of the IT system, provide user tech support.
Developer	Maintain and problem-solve operational code.

Implementation Feature Overview

We recommend a modular and extensible architecture be used in the design of an n-of-1 trial platform.[5] In this spirit, we introduce a list of desired capabilities of a trial platform (Box 5–1) as a guide to evaluating approaches for automating n-of-1 trials, or as a jumping-off point for designing a new approach.

An n-of-1 trial platform will also need to support a variety of user roles at different stages of the process with different access to information maintained by the platform (Table 5–1). Not all technical platforms need to implement all roles explicitly.

Example System: MyIBD

The MyIBD[6] IT platform was developed by the Cincinnati Children's Hospital and Medical Center (CCHMC) and a third-party consulting group (including author I.E.) as part of its Collaborative Chronic Care Network (C3N) health services research project. They targeted a minimal set of requirements to facilitate definition and management of up to 100 concurrent, independently designed n-of-1 trials. The platform is part of a personalized learning system intended to capture evidence from a patient's daily life to improve and augment patient-provider dialog. The system is intended to validate the efficacy of treatments with known heterogeneity of response, or to evaluate other treatments for which minimal research exists. Aggregation of n-of-1 trials was not a goal of the version of MyIBD reported here.

A Web form is used to capture the trial goals and design constraints. The system supports simple A-B treatment responses as well as multiphase withdraw/reversal or alternating designs. Trial outcomes are monitored using Shewhart-style statistical control charts (Figure 5–2). A single data review screen provides a scrollable view of all measures. The measures are plotted on an Xbar control chart using 3-sigma control lines calculated from the first 20 measured data points (the minimal baseline period for this platform).

The MyIBD system currently supports three roles: Administrator, Researcher, and Patient. A Clinician role linked to a subset of patients with clinician population management features is planned. The service consists of a simple administrative dashboard to create and review user accounts. Administrators are the only users able to see patient identifiers. Researchers are shown a de-identified list of the patient population (Figure 5–3) and can click to review individual charts.

Box 5–1. Requirements of n-of-1 trial platform

Features supporting n-of-1 trials

- Record clinician goals and patient goals
- Document the experimental hypothesis
- Protocol implementation support
  - ◊ Library of characterized treatments (including details of onset, carryover, etc.)
  - ◊ Library of characterized measures (including precision and variance)
  - ◊ Support for randomization
  - ◊ Web service connections to acquire/share libraries of standard measures
- Trial protocol specification
  - ◊ Choice of characterized treatments
  - ◊ Choice of measures
  - ◊ Choice of duration and number of treatment periods
  - ◊ Decision on important covariates to track
  - ◊ Analytical design
- Connection to Electronic Medical Records (EMRs), Personal Health Records (PHRs), pharmacy records (obtain medication context, lab reports, etc.)
- Data collection and user engagement support
  - ◊ Data capture modules (e.g., choice lists, visual analog scales)
  - ◊ Applications programming interfaces (APIs) to third-party data services such as sensors, apps (e.g., for symptom tracking)
  - ◊ Direct email or Short Message Service (SMS) submission of patient-reported outcomes (PRO)
  - ◊ Trial progress review screens for patients and clinicians, and other user engagement modules (e.g., leaderboards, rewards)
- Data analysis and review
  - ◊ Data preprocessing modules
  - ◊ Statistical analysis modules
  - ◊ Visualization modules
  - ◊ Data review and decision-support modules

Institutional support for n-of-1 trials

- Integration with electronic health records (EHRs) for recruiting and screening
- Configurable eligibility requirements
- Support for external informed consent processes and documentation requirements
- Population review
- Summary reports (e.g., participation, utilization)

Aggregation of n-of-1 trial results

- De-identification of patient record (for real-time in situ analysis, or for download to external systems for secondary analysis)
- Statistical analysis and aggregation of raw individual patient-level data
- Statistical analysis and aggregation of summary results data
- Statistical analysis and modeling of aggregated outcomes
- Models for using aggregated group outcomes to facilitate "borrow from strength" for individual treatment effects and to estimate individual-level heterogeneity of treatment effect

IT infrastructure

- Secure data storage
- Data transmission security
- Data downloading in multiple formats
- Authorization controls (who can do what)
- De-identified views of data

Figure 5–2. Control charts

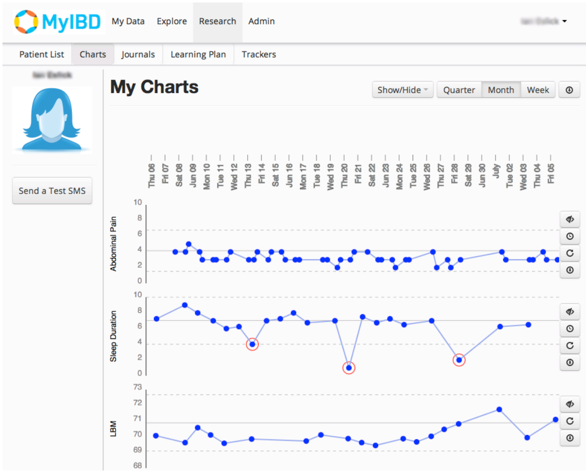
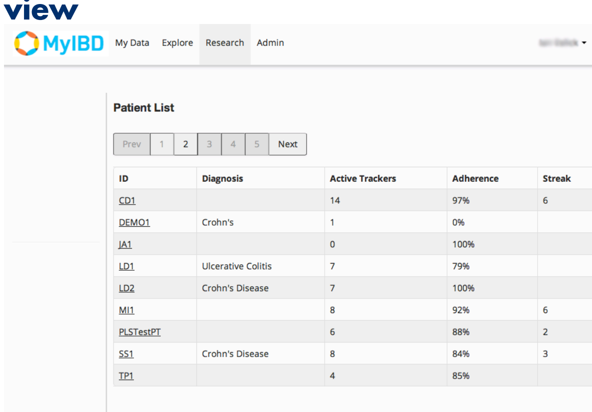


Figure 5–3. De-identified population view



ID	Diagnosis	Active Trackers	Adherence	Streak
CD1		14	97%	6
DEMO1	Crohn's	1	0%	
JA1		0	100%	
LD1	Ulcerative Colitis	7	79%	
LD2	Crohn's Disease	7	100%	
MI1		8	92%	6
PLSTestPT		6	88%	2
SS1	Crohn's Disease	8	84%	3
TP1		4	85%	

Users are given a dashboard showing outstanding data recording actions, recently recorded data, and additional reports for configuring third-party services, updating data recording schedules, and a generic journal entry facility (Figure 5–4). Due to the challenges of integrating with the hospital's commercial EMR and network registry, manual entry of medication and treatment periods is facilitated via the Journal Entry mechanism.

The initial target population is pediatric patients with IBD such as Crohn's and ulcerative colitis. An integration of standard measures is provided including PROMIS Fatigue (weekly), PROMIS Pain (weekly), and PEDS QL (monthly). The system supports a user-extensible catalog of measures, treatments, and experimental protocols (Figure 5–5). This catalog is maintained manually and reviewed by researchers, and made available to patients and clinicians. For now, no procedures have been implemented to import measures from external data sources.

Figure 5–4. User tasks

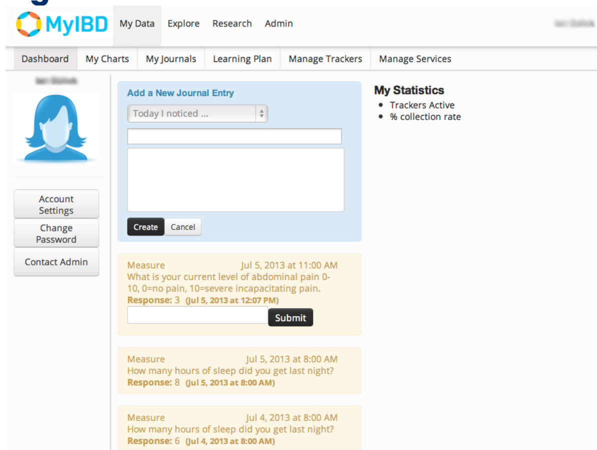
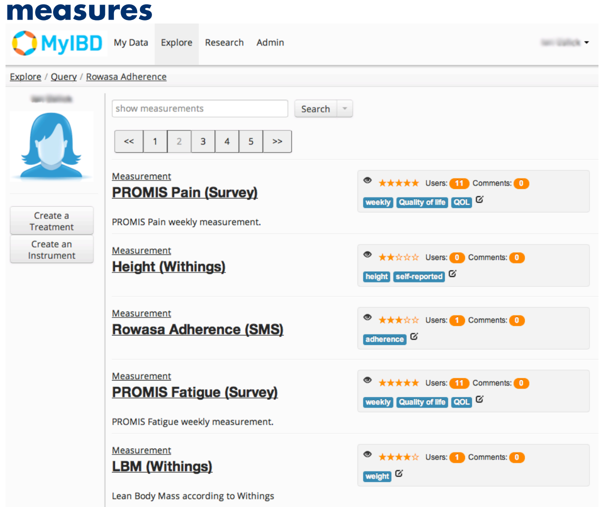


Figure 5–5. Catalog browser for measures



The clinical team had opted not to support treatment blinding at the time of this writing, as many of the planned treatments, such as diet, home-supplementation, or lifestyle changes, are difficult to blind. Consequently, all of the collected data and treatment context are available to both patients and clinicians throughout the trial.

The full development costs for this system are anticipated to be over $250,000. The ongoing infrastructure and maintenance costs are currently $400/month. This fee includes software and service licenses and a set of four cloud-hosted servers supporting high availability, backups, Short

Message Service (SMS) transaction fees, and other miscellaneous costs. Additionally, a $1,000/month support contract was in place for the first year to secure 1–2 days/month of consultant time for critical bug fixes and small feature enhancements. The project is exploring making the existing functionality more widely usable, including a software-as-a-service offering and open-source licensing of the underlying code base.

The benefit of this large up-front investment is that sustaining and per-patient costs are minimized, requiring only a small per-user fee. A single installation of the MyIBD platform will eventually scale to thousands of patients with only minor changes to the user interface. Scaling can also be accomplished through deploying the service over multiple servers. The incremental per-user cost in the current system is almost entirely driven by usage fees for the SMS gateway service. Tracking three daily values per patient requires approximately 90 messages/month. At one cent per message and with thousands of total users, the amortized IT cost per user per month is estimated to be around $1.00. While asymptotic infrastructure costs can be very low, this should not indicate that using a trial system is inexpensive. The majority of costs will be in services such as user support, technical support, ongoing development, multilingual translation, statistical and methodological review, and clinician review. Fortunately, many of these service costs are amortizable (e.g., translations and technical support).

Design Considerations

The following sections take a deeper look into many of the important components of an n-of-1 trial platform listed in Box 1 and discussed above. As introduced in Chapter 1 and demonstrated in the MyIBD example, n-of-1 trials ensure validity of the trial inference through two primary mechanisms: protocol-defined, time-varying exposure, and systematic measurement of outcomes (patient reported and others) via EMRs, Web sites, mobile devices, and sensors. Proper interpretations of these data require a data cleaning and analysis pipeline to generate findings to populate result reports and visualization. The tradeoffs involved in defining an n-of-1 protocol can be subtle and complicated. Not all providers are equally facile at measure or trial design, so we recommend the inclusion of features to import or generate libraries of measures and protocols to facilitate reuse.

Time-Varying Exposure

Time-varying exposure refers to the restriction on patient behavior during different treatment periods to ensure that he/she is exposed to the alternative treatment conditions being compared (one of which can be the control condition, including no treatment, placebo, or current treatment) according to a prespecified schedule. As discussed in Chapters 1 and 4, the schedule of exposure needs to account for the operating characteristics of the treatment conditions, such as onset time, washout (for withdraw/alternating designs), and the variability of the measurement (how many samples are needed over what duration of time to get an accurate estimate of the outcome).

For n-of-1 trials, the issues involved in treatment adherence are accentuated due to the complexity of scheduled switches between treatments being tested. Treatment adherence may be facilitated by automated reminders, generated from the IT system, of what behavior is needed at a particular point in time, as well as (for drug treatments) prepackaged dose packs for specific treatment intervals managed by the pharmacy.[7] IT systems can also support blinding by masking the clinician, pharmacist, and the patient to the patient's assigned treatment.

An IT platform should accommodate the effects of unanticipated events (such as hospitalization, vacations, nonadherence) by allowing the study protocol to be adjusted midstream. For example, data collection may need to be suspended for a period of time, or phases of the trial may need to be restarted, possibly including reverification of the patient's baseline. Changes may need to be made in trial execution (e.g., producing reminders, interacting with a pharmacy) or in analysis. This highlights the difference between an intended "planned study protocol" and the actual "executed study protocol."[8] Accomplishing automatic adjustments requires that the system maintain a representation of treatment effects such as onset and washout periods or facilitate explicit changes in the schedule by an expert (see related discussions in the section "Adaptation" in Chapter 4).

Measurement

The second critical component of an experiment is the measurement of observations of the patient over time. In published n-of-1 trials, measurements often consist of questionnaire responses or lab values at the end of a trial phase; however, n-of-1 trials are increasingly leveraging time-series data, especially with mobile technologies that enable frequent "ecological momentary assessments"[9] (EMA) of patient symptoms. Time-series data involve repeated daily or weekly measurements that are taken within a given phase of a trial (see the section "Statistical Models and Analytics" in Chapter 4). Each measurement may be a single value such as weight or a compound multidimensional assessment, such as:

- **Time point or time unit:** Length of time since start of a treatment period, or a timeframe of validity (e.g., over the last week)
- **Measure(s)** (integer, decimal, categorical, ordinal, free text, compound, complex):
    - ◊ Devices often generate multiple and compound measures for a single time point (e.g., systolic and diastolic blood pressures or GPS latitude/longitude)
    - ◊ Complex responses such as survey items[10] may be summarized into summary scores or indexes
    - ◊ Devices may need to be calibrated and harmonized, especially if a variety of different devices are used for the same measure
- **Definition:** What does this measure represent? Often this includes binding to a terminology or coding system (e.g., LNC 1558-6, the LOINC code for fasting blood glucose), which is particularly important for aggregating data across multiple trials and/or centers.

  In any trial, but especially in n-of-1 trials, the context for a data element may have a large effect on observed measures and on overall trial results. For example, if a patient records a pain score of "2 at 10:30 p.m. EST on Tuesday, March 5, 2013," we may also want to know:
    - ◊ **Mode.** How did the patient record this measure? Was it via telephone, SMS, Web application, a measurement device, or pencil and paper?
    - ◊ **Time recorded.** When did the patient record the measure? He or she may have forgotten to record that morning and instead recorded it that evening, for example, reporting at 10 p.m. a pain score that was actually experienced at 2 p.m.
    - ◊ **Prompt.** What, if any, prompt elicited the measure?
    - ◊ **Schedule.** What was the scheduling on this prompt: momentary assessment, or scheduled? Randomization of prompts may improve the accuracy of the resulting data.
    - ◊ **History.** Were the data ever changed or updated?
    - ◊ **Respondent.** Were the data entered by the patient, entered by a proxy (e.g., parent, caretaker), or recorded by an automated device?

These "data about data" are called metadata. It is critical that a trial IT platform has robust support for collecting, storing, and analyzing metadata, because these contextual factors can interact with time-series data and greatly impact trial inference. For trials with small effects or poor precision (e.g., because the number of treatment periods is limited), these interaction effects may overwhelm and mask the underlying effect of interest. Metadata are particularly critical to facilitate aggregation of data from individual trials to estimate population effects or to predict the likely outcome of future n-of-1 trials. In the absence of existing standards around metadata capture, a platform should support an open-ended set of elements and allow providers of observations and other data to expand the set of labels over time without requiring central coordination or standards.

Statistical Analysis Modules

Chapter 4 identifies a variety of options for performing statistical analysis of n-of-1 trials, including simple statistical tests on observed results (t-test, ANOVA, etc.), regression analysis, and Bayesian analysis using closed-form solutions or numerical solutions such as techniques based on Markov-chain Monte Carlo. A variety of statistical procedures are needed for model selection, filling in missing data, adjusting for time-series effects (e.g., autocorrelation models), removing short-term special-cause variations, preprocessing the data for visualization, et cetera. Over time, different measures and different trial conditions may require new forms of analysis.

It is essential that the IT system supports the automation of the statistical procedures needed (see the section "Automation of Statistical Modeling and Analysis Procedures" in Chapter 4). For statistical procedures that are implemented in a statistical language such as R or Matlab, export procedures for analyzing the data in external packages are often needed. R, for example, can be embedded or linked as a Web service to other software, which may simplify the creation and extension of the statistical facilities of the trial platform. Extension of the statistical facilities may become especially important if the platform incorporates in situ aggregation techniques, already an active area of research in statistics and a likely active area of research and development in IT in the coming years.

Visualization Modules

For n-of-1 trials with strong effects, it might be straightforward to analyze the trial outcomes using simple visual techniques. Visualizations typically involve a data preparation phase (data transformation, filtering, etc.), a customization of the data presentation (colors, emphasis, and labeling), and a rendering phase. If the trial platform is Web based, there are many off-the-shelf charting packages that may suffice, but for implementing the annotations that are often helpful to interpreting n-of-1 trial results, a more flexible graphics language such as D3 may be a more appropriate foundation.

Many clinicians and most patients have limited statistical numeracy.[11] N-of-1 trials therefore require very clear communication of trial results that use visual heuristics and user-friendly, comprehensible statistical interpretation of the trial findings (see the section "Presentation of Results" in Chapter 4).

Health IT systems are typically designed for clinical settings and often lack an effective interaction design[12] (i.e., the workflow and experience of a software-enabled process that encompasses the user interface). Designing and implementing a user interface is complex and costly, but building a seamless and engaging user experience can be an order of magnitude more challenging and expensive. With increasingly sophisticated user experiences becoming commonplace in consumer software (e.g., Apple iPhone), many patients and clinicians will expect equally sophisticated design from health IT. Thus, IT platforms designed to scale up n-of-1 trials must devote adequate resources to user interaction design to ensure user uptake and engagement.

Sharing

Important capabilities of n-of-1 platforms include sharing of n-of-1 outcomes among providers and researchers for secondary analysis, aggregation of results for estimation of population effect, and population management. Platforms should export both identified and de-identified datasets, or support dynamic updating of population-oriented data systems such as i2b2[13] or SHRINE.[14]

Further, the move to patient-centered care has increased the emphasis on enabling patients to easily access their own data. An emerging consumer ecosystem is using longitudinal health data for exercise performance, weight-loss coaching, personal health tracking, Quantified Self,[15] and other activities. Making data available via an applications programming interface (API) to third parties, with explicit patient consent, will be an important capability of future systems. The Open mHealth project is a leading effort to provide interoperability standards at the data and protocol layers.[16]

Another form of sharing that can be valuable is for patients to share their own data, displays, and reports with family, friends, peers, or over social media channels. This level of semipublic sharing may not be suitable for all provider settings, but facilitating patient use of these sources of social support and reinforcement can have therapeutic value. Sharing data over social media can be enabled through exposing patient-approved, public visualizations of the data as sharable URLs or through social media via n-of-1 platforms that send status updates.

Templates and Libraries

As described above, a library of treatments and measures will help to simplify the process of trial design and specification, to increase methodological strength, and to enhance user engagement. Designers of n-of-1 trials can consult such a library to find detailed information about treatments, including their speed of onset, washout periods, and other necessary information in designing a proper n-of-1 trial. Detailed information on outcome measures would also be helpful. Measurements have statistical properties that can be characterized for the average respondent, such as reliability, variance, and reproducibility. Measurements may also be subject to biases, including practice effects, onset behavior, etc. These characteristics of treatments and measures must be taken into account to ensure methodologically strong scheduling and analysis of a specific trial. Examples of preexisting libraries for measures include: PROMIS,[10] GEM,[17] NeuroQOL,[18] NIH Toolbox,[19] and PatientsLikeMe.[20]

An n-of-1 methodology library should also include parameterized templates of successful trials that can be used or adapted "off the shelf" by other patients or providers on the same platform with similar study questions. This will reduce the barrier for clinicians and patients who do not have the statistical or methodological expertise to design and run n-of-1 trials on their own. Classes of trials that have been successfully reviewed by statisticians and methodologists may be fully automatable, reducing overall personnel costs and burden, while increasing the methodological quality of executed studies.

A methodology library such as we discuss here could be a component of a full-featured n-of-1 IT platform, or it could be a common shared resource. The more closely such a library is integrated into an operational IT platform, the more easily those treatment and outcome measure characteristics could be populated directly from the results of prior trials (e.g., the typical within-subject and across-subject variations of a measurement).

Cross-Cutting Concerns

Accessibility

Technology promises to help close health disparities, but underserved populations have special needs that require special attention. The user interaction design should be culturally sensitive, and instructions and prompts should be adapted and translated expertly to accommodate multicultural, multilingual populations. Section 508 compliance[21] should also be sought if the targeted audience includes those visually or hearing impaired.

Privacy

An n-of-1 trial platform needs to be designed with the same consideration given to any health IT system that maintains patient data. The goal is to facilitate patient access, clinician utility, and third-party review with minimal effort, while preserving privacy consonant with ethical principles and applicable regulations. De-identification of data can provide privacy protection sufficient to enable authorized people to review records and explore aggregate data analysis safely and ethically. However, some types of data (e.g., location traces, genomic data) are almost impossible to de-identify. Privacy is best maintained and assured by a combination of technology and policy.

In U.S.-based settings where research is performed on the data collected by the platform, all developers and system administrators will need to complete Human Subjects Training, as they will have physical access to patient identifiers. (See Chapter 2 for a more complete treatment of human subjects issues relevant to n-of-1 trials.)

Data Security and HIPAA/HITECH

U.S.-based systems that store or operate on patient data must adhere to a set of regulations created by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and Health Information Technology for Economic and Clinical Health Act (HITECH). HIPAA and HITECH present the rules that govern the security and privacy of Protected Health Information (PHI), such as names, dates of service, and contact information that are provided to "covered entities," Such as health plans for health providers. These provisions are relevant to platform and service providers because covered entities are required to ensure that they have formal Business Associate Agreements (BAA) with any vendor that processes patient information. The vendor, in turn, must have a BAA with any of their vendors, such as cloud service providers, that store or transmit PHI in nonencrypted form, even if such access is transient and limited.

Some of the provisions of HIPAA require that PHI be encrypted both "at rest" and "in transit." This means that data stored on the platform servers or in backup archives must be encrypted, and data transmitted between components of the platform (such as between Web server and database, or between Web servers and clients such as Web browsers) must also be encrypted. Further, all procedures for managing access to data and system administration must be formally documented and controlled. There are two approaches to encryption: bulk encryption, such as an encrypted database or file system, and "column" encryption, singling out HIPAA-delineated identifiers and encrypting them at the application layer. Unless there is explicit database support for column encryption, it can increase the complexity of applications to explicitly manage encryption and decryption of specific fields. Indexing and searching over PHI fields are often required in IT systems and can be facilitated by creating proxy fields that are one-way hashes (enabling indexed lookup of patient email) or transformed values (such as date of birth truncated to a year) that allow a query to restrict to a superset of the desired range of dates.

One of the most challenging aspects of adhering to the latest regulations in the emerging ecosystem of data sources is identifying the line of demarcation between a trial platform and a third-party service. Standard SMS, for example, is not a secure, HIPAA-compliant communications channel, and the third-party SMS provider is storing the patient's mobile phone number, a piece of PHI. However, regular text messaging is far easier for patients to use than "secure" text messaging solutions developed for provider-to-provider communication. The same rules apply for consumer devices and services such as Fitbit[22] and FitnessKeeper[23] that collect and store consumer data on systems that are typically not HIPAA compliant.

The intent of HIPAA was to protect patients and improve their access to data, even if in practice it appears to make access more difficult. Under the 2013 omnibus regulations,[24] communication channels such as SMS and email are acceptable for patient-provider communication and data exchange if users have explicitly asked to exchange data over a particular channel, the risks are modest, and they have been informed of those risks.

It is important to designate someone in the provider organization, even if outsourcing to a third party, who can review the organization's obligations under current legislation with regard to risk exposure, documentation, and notification in the event of a breach. That person should audit the platform procedures and recommend improvements and/or fixes on a regular basis and prior to major updates.

Authentication and Authorization

One critical consideration for platforms that provide interfaces to users (clinicians, patients, or any other user) is how those users will be authenticated to the system. Authentication is the process of determining whether the user is indeed the intended user, and authorization is the logic that determines what a specific user can do. N-of-1 systems are particularly amenable to role-based authorization, where a given user satisfies one or more roles that in turn dictate what data or reports they have access to. For example, clinicians can review only data on their own patients, and reviewers can see only de-identified population-level data and review summary records.

Everyone faces the problem of having to remember passwords for a wide variety of systems. Where possible, an n-of-1 platform should seek to integrate with existing authentication systems to facilitate easy credential recovery according to industry best practices. At the time of this writing, the OAuth[25] standard is emerging as the most widely used consumer authorization framework for third-party access to data via APIs. Within the health care sector, discussions around a national Unique Patient Identifier involve many technical and policy challenges.[26] Current thinking is to ensure unique identification through a combination of patient identifiers and processes (e.g., two-factor identification), and any n-of-1 platform should align with the identification policies specific to health care.

Extensibility

Though n-of-1 trials are an old technique, they remain novel in most health settings and have received limited attention from the academic research community. Techniques and trial design styles are likely to evolve, and it is prudent for trial platform designers to adopt a modular and extensible approach to facilitate the adoption of new techniques over time. Ideally, extensibility is made possible through a "plug-in" architecture, allowing third parties to add functionality to a well-defined API or data format without requiring a deeper understanding and extensive remodeling of the existing platform.

The areas most important to emphasize extensibility include: a catalog for reusable components and templates (e.g., standardized measures like PROMIS), user interaction templates for surveys (to facilitate development and adoption of new adaptive assessment models), processing modules, statistical modeling and analysis models, visualization modules, and shared decision support.

Open mHealth[16] defines an open data and software interoperability approach that is congruent with this perspective along with a specification for measures. It defines a framework for modular assembly of data processing, statistical modeling and analysis, and visualization modules to create specific time-based or summary views of time-series data, and allows integration of Web-based services. For example, Web services can be provided to match a patient's electronic health record data to heuristics on types of patients and situations amenable to n-of-1 studies (discussed in Chapter 6), and thus suggest to the clinician to consider n-of-1 studies for the specific patient. However, the state of the art in computational eligibility determination is quite rudimentary.[27]

Another model to consider is the Substitutable Medical Apps Reusable Technologies (SMART) platform[28] a standard that defines a model for pluggable applications for EMR systems. SMART defines a Web-based model for interoperability, allowing third-party Web-based user interfaces to plug into the patient portal of an EMR or personal health record. A SMART approach could support the provision of n-of-1 studies as one of several point-of-care research studies that clinicians can "order."[29] Providing similar facilities would allow new developers to build shared decision support screens, Web-based instrumentation, or new visualization and/or processing solutions for a core platform.

Platform Economics

When considering buying or building an n-of-1 platform, the total cost of ownership should be carefully analyzed. Though initial development of a trial platform may be a fixed cost and born by a research grant or one-time funds, ongoing support costs, particularly as needs evolve, can be substantial. Resources will be required for ongoing user support, technical support, hosting costs, and new feature development. Adaptation and translation costs for multicultural, multilingual support can be significant not only up front, but also over time as site content is updated. Service costs can also be significant if users are contracting with third-party services such as a gateway for eliciting data via SMS, telecom costs for phone/fax, and transcription costs for manually transcribing paper responses. Institutional owners of n-of-1 platforms may consider recharge mechanisms for defraying carrying costs.

Summary

Development of a custom IT-based n-of-1 trial platform calls for significant up-front investment, as illustrated by the MyIBD example. Moreover, it requires significant ongoing investment for both clinical and IT operations. Given the high costs, the lack of strong evidence establishing value, and the small market, there are currently no commercially available n-of-1 trial platforms. It is likely that such platforms will be developed instead with government or foundation funding intended to characterize the applicability and use of n-of-1 trials at scale.

Institutions interested in IT support for n-of-1 trials will find it prudent to maximize knowledge transfer and to amortize investments across multiple institutions by leveraging existing open-source projects and approaches. If inhouse development and management of a clinical trials platform is impractical, it may be possible to use one of these open-source platforms through collaborations with other institutions hosting such platforms.

Another option is to investigate reuse or extension of existing clinical trial management or other data acquisition systems, although many clinical trial management systems are designed explicitly for traditional parallel group randomized trials and might be difficult to adapt to the time-varying exposure design in n-of-1 trials. In time, commercial services and/or software offerings may lower costs of ownership and increase functionality beyond what is offered by open-source alternatives.

Nevertheless, it is possible to facilitate many n-of-1 trial activities without the comprehensive design approach advocated here. MyIBD provides one example of simplifying and accelerating n-of-1 trial deployment with only a small subset of these features. However, each of the features advocated here will expand the population a platform can serve with greater ease of use and reduced costs.

Checklist: N-of-1 Trial IT Platform Selection and Deployment Strategy

Guidance	Key Considerations	Check
Determine the purpose for which you are creating or buying an n-of-1 trial IT platform	• Recommendations depend on purpose. • Common purposes include: improving care delivery, evaluation of benefit, enabling health services research, evaluating specific methodologies, generating reusable knowledge, integration with a larger learning system.	❒
Decide whether to build or buy	• Perform an assessment of lifetime costs. • Account for costs of training, education, user support, statistical consultant, and technical support. • Account for future needs and access to developers if developing or using open-source offerings. • Build only if you are performing research on trial design, statistics, or implementation and have access to a captive development team. • Buy if you are able. • Avoid using third-party contractors on one-time research-funded contracts.	❒
Decide on open- or closed-source solutions	• Use open-source if you plan to innovate on the platform itself. • Prefer open-source, but choose the best solution if your goals include improving clinical care delivery.	❒
Choose a hosting model. Is the service hosted in your institution's facility or managed by external resources on servers not under your direct control?	• A cloud solution is preferred if it satisfies your institution's HIPAA and/or Institutional Review Board obligations and integrates with your clinical systems where needed.	❒
Define patient ownership of and access to data	• Patients should have direct access to their raw data after a trial is completed. • Ideally, provide a patient portal for access during and after the trial with user-appropriate reports and visualizations. • Support data download using standards such as Open mHealth, BlueButton,[30] and/or simple comma-separated values (CSV). • Provide an API to enable third-party services to pull data on behalf of users, for example, via OAuth.	❒

Checklist: Feature Requirements of an N-of-1 Trial IT Platform

Guidance	Key Considerations	Check
Assure that platform is sufficiently flexible to support the range of anticipated n-of-1 designs	• Involve methodologists and statisticians in developing the design specifications of the system.	❒
Protocol design support	• Support a catalog of treatments, measures, and experiments. • Support user interface workflow to create new measures, treatments, and experiments. • Ensure the platform supports all likely trial designs (randomized, counterbalanced, blocked, adaptive, sequential stopping, etc.). • Platform should allow for designs to be extended over time (via a modular design).	❒
Provide a population management view	• Support de-identified access to trial cases for statistical review. • Support configurable blinding of investigator and patient accounts.	❒
Adaptive schedule management	• Enable users/clinicians to restart trial phases, annotate special causes, etc. • Allow for patient-driven selection of data collection prompts and/or reminders.	❒
Provide a Web-based portal for trial review by all participants	• Provide a portal for review of all filtered trials, including summaries of progress, adherence, and any electronic conversations. • Provide integrated methods for patient contact. • Provide visualizations of outcome data after any blinding periods have expired.	❒
Provide built-in data collection facilities	• Provide built-in assessment tools for common measures available via prompts, Web survey tools, email, and paper. • Provide standard instruments, where possible.	❒
Support download of trial data for post-trial analysis	• Allow for de-identified download of raw data for additional statistical review in case platform analysis is insufficient for a specific trial.	❒
Obtain requisite data from the electronic medical record	• Integrate with the medical record to populate contextual information, including medication and demographic information. • Provide automatic access to lab results. • Optional: Provide support for manual entry and display of this information alongside trial results.	❒
Enable connection to pharmacist services	• For platforms testing drugs, provide support for printed, e-fax, and email of pharmacy instructions. • Instructions should include support for randomization schedules, blinding, and placebos.	❒

Checklist: Optional Features of an N-of-1 Trial Platform

Guidance	Key Considerations	Check
Provide multilingual and culturally sensitive versions	• Must support if deployment is anticipated with multilingual and/or culturally diverse populations. • Translations of common measures, reminder prompts, etc., should be shared in common libraries.	❒
Ensure Section 508 compliance if applicable	• Depending on the anticipated patient population, accommodation for patients and clinicians with auditory, visual, and physical disabilities is needed. • Government agencies are subject to Section 508 (29 USC 794d).	❒
Integrate with other institutional IT systems	• Support external authentication schemes to reuse existing credentials, for example, from patient portals. • Optional: Embed user interface (via iframe tag) into institutional portals or intranets.	❒
Provide printed forms and reports; support manual transcription from paper	• Expand the reach of the platform to underserved populations and populations without connectivity. • Enable consistent data capture when electronic systems are unavailable or inaccessible for any reason.	❒
Support scanning of printed records	• Optional: Support for scanning and/or OCR (optical character recognition) of paper records will reduce workflow costs and enable de-identified transcription.	❒
Interoperate with third-party services	• Provide support for importing data from mobile devices, medical and consumer devices, and third-party service platforms.	❒

Checklist: Additional Considerations for an N-of-1 Trial Service

Guidance	Key Considerations	Check
Provide educational materials	• Platforms should provide support for educational materials and aids embedded in the IT infrastructure. • Provide support for developing and maintaining culturally relevant translations of educational content as required to serve the target population.	❐
Simplify human subjects research	• Support e-consent procedures. • Support unique ID generation or import of unique IDs generated elsewhere.	❐
Simplify methodology review	• Provide facilities for online and offline methodology review.	❐
Provide user support	• Ensure that a nurse practitioner or the equivalent with n-of-1 trial experience can respond to questions such as what to do about missed treatment, lost data entry, and medication side effects. • Optional: Provide an integrated live chat feature in the patient portal.	❐
Provide technical support	• Provide a telephone number to call for technical support and email address with turnaround guarantee. • Technical support should include a formal issue tracking system.	❐

References

1. Scuffham PA, Nikles J, Mitchell GK, et al. Using N-of-1 Trials to Improve Patient Management and Save Costs. Journal of General Internal Medicine. 2010;25(9):906-913.
2. Larson EB. N-of-1 Trials: A New Future? Journal of General Internal Medicine. 2010;25(9):891-892.
3. Larson EB, Ellsworth AJ, Oas J. Randomized Clinical-Trials in Single Patients during a 2-Year Period. Journal of the American Medical Association. 1993;270(22):2708-2712.
4. Crandall W, Kappelman MD, Colletti RB, et al. ImproveCareNow: the development of a pediatric IBD improvement network. Inflamm Bowel Dis. 2011;17:450-457.
5. Chen CN, Haddad D, Selsky J, et al. Making Sense of Mobile Health Data: An Open Architecture to Improve Individual- and Population-Level Health. Journal of Medical Internet Research. Jul-Aug 2012;14(4):126-134.
6. Eslick I, Kaplan H, Chaffins J, et al. MyIBD. 2012; Accessed July 14, 2013. https://myibd.c3nproject.org.
7. Guyatt G, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. CMAJ. Sep 15 1988;139(6):497-503.
8. Sim I, Niland JC. Eligibility and Protocol Representation. in Richesson RL, Andrews JE (Eds.) Clinical Research Informatics. February, 2012; London: Springer.
9. Shiffman S, Stone A, Hufford MR. Ecological momentary assessment. Annual Review of Clinical Psychology. 2008;4:1-32.
10. The National Institute of Health. The Patient-Reported Outcomes Measurement Information System (PROMIS), 2013. Accessed July 14, 2013. http://nihpromis.org.
11. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, et al. Helping Doctors and Patients Make Sense of Health Statistics. Psychological Science in the Public Interest. 2008:53-96.
12. Cooper A, Reimann R, Cronin D. About face 3: the essentials of interaction design. 2012; Indianapolis, IN: Wiley.

13. Kohane IS, Altman RA. Health-information altruists—a potentially critical resource. New England Journal of Medicine. 2005;353:2074-2077.

14. Weber GM, Murphy SN, McMurry AJ, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. Journal of American Medical Informatics Association. 2009 Sep-Oct;16(5):624-630.

15. Wolf G. Know Thyself: Tracking Every Facet of Life, from Sleep to Mood to Pain. Wired Magazine. 2009:24(7):365.

16. Chen C, Haddad D, Selsky J, et al. Making sense of mobile health data: an open architecture to improve individual- and population-level health. J Med Internet Res. 2012;14(4):e112. doi: 10.2196/jmir.2152.

17. National Cancer Institute, GEM Grid-Enabled Measures Database. Accessed July 14, 2013; www.gem-beta.org.

18. Northwestern University, NeuroQOL: Quality of Life in Neurological Disorders, 2013.Accessed July 14, 2013. www.neuroqol.org.

19. National Institutes of Health and Northwestern University, NIH Toolbox: For the assessment of Neurological and Behavioral Function, 2006-2013. Accessed July 14, 2013. www.nihtoolbox.org.

20. Frost, J, Massagli, M. Social uses of personal health information within PatientsLikeMe, an online patient community. What can happen when patients have access to one another's data. Journal of Medical Internet Research. 2008;10(3):e15.

21. Office of Federal Contract Compliance Programs, Section 508 of the Rehabilitation Act of 1973, as amended. 29 USC Section 793. 1993; www.section508.gov.

22. Fitbit, Incorporated. 2013; www.fitbit.com/company.

23. Fitness Keeper, Inc. 2013; http://runkeeper.com.

24. Federal Register Vol 78, No 17, 45 CFR Parts 160 and 164. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules Under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; Other Modifications to the HIPAA Rules 2013 Department of Health and Human Services.

25. Internet Engineering Task Force: The Oauth 2.0 Authorization Framework. RFC 6749, 2012; http://tools.ietf.org/html/rfc6749.

26. Analysis of Unique Patient Identifier Options. National Committee on Vital and Health Statistics. 1997. www.ncvhs.hhs.gov/app3.htm.

27. Cuggia M, Besana P, Glasspool D. Comparing semi-automatic systems for recruitment of patients to clinical trials. International Journal of Medical Informatics. 2011 ;80(6):371-388.

28. Mandl KD, Mandel JC, Murphy SN, et al. The SMART Platform: early experience enabling substitutable applications for electronic health records. Journal of the American Medical Informatics Association. Jul 2012;19(4):597-603.

29. Fiore LD, Brophy M, Ferguson RE, et al. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. Clin Trials. 2011 Apr;8(2):183-195.

30. Office of the National Coordinator for Health Information Technology, HealthIT.gov. 2013. Accessed July 14, 2013. www.healthit.gov/bluebutton.

Chapter 6. User Engagement, Training, and Support for Conducting N-of-1 Trials

Heather C. Kaplan, M.D., M.S.C.E.
Cincinnati Children's Hospital Medical Center

Nicole B. Gabler, Ph.D., M.P.H., M.H.A.
University of Pennsylvania

The DEcIDE Methods Center N-of-1 Guidance Panel*

Introduction

Traditional research approaches do not always provide the type of personalized evidence necessary for patients to make fully informed decisions. In the absence of individualizable evidence, defined as evidence that is directly applicable to an individual patient, clinicians and patients use imperfect approaches to personalize care. Since it is often difficult to apply evidence from randomized controlled trials (RCTs) directly to the care of individual patients, clinicians use clinical judgment to decide whether trial results are applicable to an individual patient. When treating chronic or intermittently symptomatic conditions, they may conduct informal trials of therapy whereby a clinician will try an intervention for a particular patient to "see if it works."[1,2] Many interventions of interest to patients, such as those that facilitate the management of side effects or those addressing lifestyle changes, have never been tested by RCTs, making evidence-based decisions among these alternatives impossible. Accordingly, patients often use self-experimentation, whereby various interventions are tested in an ad hoc fashion without any underlying scientific rigor.[3,4] It is likely that the informal approaches being used by clinicians and patients to personalize care are flawed and can be improved.[5]

N-of-1 trials are a useful method to personalize care but are not widely used (see Chapter 1). While n-of-1 trials were introduced to the medical community in the late 1980s and early 1990s, inspiring the creation of n-of-1 trial services at several academic medical centers,[6,7] the burden of conducting the trials eventually led to a drop-off in popularity. Recently, however, there has been a resurgence of interest in n-of-1 trials, both independent of and in cooperation with n-of-1 consult services.[2,8] However, this resurgence may be doomed to fail if specific attention is not paid to engaging, training, and supporting patients and providers interested in conducting n-of-1 trials. This includes addressing problems encountered by both patients and clinicians as well as designing systems that support training and execution of trials with fewer burdens.

This chapter will discuss methods of increasing patient and provider engagement, training, and support, with the aim of presenting a framework that will facilitate the ease and conduct of n-of-1 trials in a sustainable way.

Patient Users

Eliciting interest, understanding, and cooperation from patients is of utmost importance when conducting n-of-1 trials. According to Guyatt,[9] "an n-of-1 RCT is indicated only when patients can fully understand the experiment and are enthusiastic about participating." Generating buy-in from patients for n-of-1 trial participation occurs both through initial engagement (approaching the patient and demonstrating the benefits of the trial) and through ongoing training and support (helping patients to understand the disease, trial procedures, data collection, analysis, and decisionmaking).

*Please see author list in the back of this User's Guide for a full listing of panel members and affiliations.

Initial Engagement and Motivation

When approaching patients to discuss initial participation in an n-of-1 trial, it is important to recognize that patients who are good candidates for an n-of-1 trial are those whose disease is easily monitored, those with disease that has been unresponsive to standard therapy, and/or those patients who are quick to respond to treatment (please also see the section "Indications, Contraindications, and Limitations" in Chapter 1).[10] It is not important that the patient meet traditional inclusion criteria for a parallel group trial; patients who are older, have comorbidities, or have lower levels of education or income are all potential candidates for n-of-1 studies.[11] However, some patient characteristics have been shown to be associated with successful trial participation, including a positive attitude, high level of motivation, intact cognitive capacity, and willingness to be proactive following poor treatment results.[10] In addition, patients who are responsible and open to both novel therapies and experimentation make ideal candidates.[10] Patients must be willing to undergo multiple treatment periods and, if important to trial design, to take a blinded medication.[6] Finally, patient recruitment works best when there is a strong and trusting relationship between the provider and patient.[10] Some patients may be unwilling to participate in an n-of-1 trial if they believe the results may lead to a recommendation to discontinue a medication they believe is effective, and this should also be addressed at the time of recruitment.[12] It is important that patients have a genuine uncertainty (equipoise) regarding which treatment is superior and a willingness to use the information from the n-of-1 trial to determine future treatment. Patients who do not express this willingness may not be suitable for an n-of-1 trial.

Partnering with patients and agreeing that the benefits of the n-of-1 trial outweigh the burdens are crucial. The following benefits of the n-of-1 trial should be highlighted when engaging patients:

1. **Personal gain:** Patients who participate in an n-of-1 trial will learn more about their own disease process and treatment than patients who receive usual care or patients who participate in parallel group trials. When patients participate in n-of-1 trials they undergo a rigorous and personalized process utilizing careful monitoring and frequent outcome reporting that can offer unique insight into their disease process[10] and patterns of symptoms related to daily activities.[13] Finally, patients will learn which treatment or drug dose maximizes benefits while minimizing adverse side effects (recognizing that treatments may affect individuals differently). Furthermore, they will be able to apply that information immediately toward their own treatment decisions, and also contribute that information to the general pool of scientific knowledge about that disease, potentially helping others.[13]
2. **Flexibility:** Unlike traditional parallel or crossover group trials, n-of-1 trials are tailored to the individual circumstances of the patient. First, the intervention tested using n-of-1 methods can be individualized for a particular patient. For trials of medication effectiveness, patients often receive individualized dosing regimens.[8] In addition, n-of-1 methods can be used to test interventions tailored to the patient's interests outside of traditional medication effectiveness, including trials of behavioral therapy or complementary and alternative medicine (CAM).[13] Because they have the potential for more side effects than behavioral or CAM trials, drug trials might run into more resistance from patients;[13] however, the n-of-1 method can be used effectively to study both medications and lifestyle modifications, offering maximal flexibility. Second, there is flexibility in the ways outcomes are selected and measured. Patients participating in n-of-1 trials can choose the outcomes that are most important to them, and the manner in which an outcome is assessed.[6] Guyatt recommends measuring a patient's symptoms or quality of life directly,[9] and all outcomes should be patient centered.[10] Data can be collected via self-administered questionnaires with Likert scales, daily diaries, and various

other formats (see Chapters 1, 4, and 5). In particular, daily diaries have been shown to be highly successful with infrequent missing data.[13]

3. **Low risk to participation:** N-of-1 trials pose low risks to patients, since the patients are already likely to be familiar with the treatments from prior use.[13] Furthermore, patients can withdraw at any point if they feel the trial is clearly ineffective or not beneficial, or if the treatment leads to undesirable side effects.
4. **Increased collaboration between providers and patients:** Patients participating in n-of-1 trials require increased monitoring and may have more appointments than other patients. This may lead to increased communication between patient and provider, ultimately better supporting shared decisionmaking.[10] This heightened communication may not only increase adherence[10] but also afford the patient greater autonomy than is typically observed in clinical practice.
5. **N-of-1 trials have been successful in the past:** Prior consult services have reported between 62 percent[14] and 84 percent[7] of trials providing a definite clinical answer, and 79 percent of patients participating in n-of-1 trials considered them useful.[6] Between 44 percent and 65 percent of patients[8,14,15] reported treatment change as a result of the trial, with between 84 percent and 100 percent[11,14,15] continuing with therapy consistent with definitive n-of-1 trial results. However, the high level of treatment continuation may not be universal across all patients and treatments.

Information about the public's motivation and interest in n-of-1 trials has been gleaned from a small number of peer-reviewed publications that conducted interviews and focus groups with patients.

Further exploration of social marketing methods, patient focus groups, and patient communities that are engaged in e-science (e.g., PatientsLikeMe, CureTogether, and DIY Genomics) may assist researchers in designing better strategies and approaches to engage patients.

In addition to approaching patients through providers, direct social marketing may have a useful role in patient recruitment. Patients outside the clinic may be interested in n-of-1 trials for many of the same reasons as patients in a clinical setting: there is value in knowing that the benefits of a particular medicine are "worth it" compared with the costs and side effects. This knowledge becomes even more valuable as patients become responsible for greater and greater portions of drug costs out of pocket. If combined with other campaigns (such as health literacy), this kind of social marketing could be even more successful than going through individual physicians.[1] In Australia, Nikles and colleagues have shown the benefits of utilizing a central administrative support structure to reach an entire country.[15] They used mainly print, TV, and radio campaigns for reaching patients, although other possibilities could include support groups, brochures in doctors' waiting rooms, and Web sites.[8] In the Australian studies, patients were able to contact the consult service directly and then provide their physicians with trial materials, including packets of medications and instructions.[15,16] It is also possible for patients to participate in n-of-1 trials completely independent of providers, although we recommend this only when the n-of-1 trial does not involve a prescription drug. In all instances, patients should discuss the results of the n-of-1 trial with their physician.

Ongoing Training and Support

Because most patients will not be familiar with the n-of-1 trial approach, researchers hoping to engage and support patients must offer education on all aspects of such trials. First and foremost, patients should have a basic understanding regarding their disease or condition as part of good clinical care. Patients will also need education and training in n-of-1 study design, especially since decisions regarding therapies, outcomes, and stopping rules are often made collaboratively among researchers, providers, and patients.[10] In patient interviews, Brookes et al.[13] found that patients see n-of-1 trials as being similar to the self-experimentation that occurs in everyday life. One patient commented, "Well yeah, well it's just like anything isn't it? You try cabbage, you don't like that, so you try broccoli."[13] Other patients found n-of-1 trials similar to conventional RCTs, but preferred n-of-1, since they would be more likely to receive two active treatments as opposed to a placebo.[10,13] Furthermore, emphasizing that variation exists between individuals and that making comparisons against oneself as opposed to others is more likely to result in a true answer for an individual patient will further differentiate the n-of-1 from more traditional trial designs.[13]

Patients are also likely to have concerns regarding the potential hazards or consequences of the trial, and these should be addressed when discussing study design. Specifically, patients may be worried about adverse drug interactions, possible suboptimal treatment for some period of time, and whether the medication will be prepared in a safe and efficient manner.[10] Finally, researcher, provider, and patient should discuss the interpretation of results prior to beginning the trial. Patients may have a desire to continue a medication despite unclear trial results or results clearly in favor of an alternative medication. Options regarding treatment after trial completion should include treatment cessation, treatment continuation, and (when the trial results are unclear) the extension of the n-of-1 trial to include more crossover periods in order to minimize uncertainty.[17]

As previously stated, one of the most important advantages of n-of-1 trials is that patients have the ability to tailor the trial to their needs and can address the outcomes (e.g., symptoms or predictors of future health) that are most important to them. Patients must be supported in identifying and defining outcome measures that clearly determine effectiveness of an intervention. Patients may be interested in exploring outcomes that have been established as effective for other patients, or they may prefer to explore outcomes on which there is little information in the literature. Regardless, researchers must emphasize the importance of consistency and accuracy in data reporting irrespective of the type of outcome measure used. As discussed in Chapters 1, 4, and 5, possible outcome measures may include quality-of-life assessments, symptom diaries, or objective outcomes (such as blood pressure measures). It is recommended that any measurement assessment be brief and easy; prior research has shown success with daily diaries.[13] Recording these outcomes can be facilitated through Web sites and other electronic technology. Some examples include using smartphones, tablet computers, and personal digital assistants (PDAs) to collect and transmit the data (mHealth), or utilizing devices and sensors (e.g., a cell phone's accelerometer) to capture data passively.[18] The role of these devices and other technology is discussed in more detail in Chapter 5. Patients must also be supported in identifying appropriate study questions. For example, it is possible that patients may prefer trials of medical devices as opposed to drug trials due to potential side effects and discomfort.[13] Furthermore, patients are unlikely to agree to participate in a lot of experiments that do not show the clear clinical benefit of one treatment versus another, so it is important to work with patients to identify possible interventions based on existing scientific evidence, the individual's prior treatment history, knowledge gained from other individual n-of-1 trials, and even anecdotal experience.

Patients will need support and training in completing the n-of-1 trial and interpreting the results. It is possible that providing support via ongoing phone, Short Message Service (SMS), and/or email from the provider or researcher may minimize dropouts and encourage adherence to data collection and trial protocols. In discussing potential trial results, the difference between clinical significance and statistical significance should be explained. As discussed in Chapter 4, significance testing might be less pertinent for n-of-1 trials; instead, the statistical methods for n-of-1 trials should provide the decisionmaker with all necessary information in a format that facilitates decisionmaking. Some experienced researchers[6,7,19] advocate using an a priori difference in outcomes as indication of clinical significance, in lieu of the exclusive use of statistical significance for decisionmaking. For example, they considered a 0.5 mean difference on a 5-point Likert scale to indicate effectiveness, since that corresponds to a meaningful improvement in well-being.[20] While we agree with the principle of emphasizing clinical significance in decisionmaking, we would like to note that it is also important to take uncertainty into consideration, to ensure that the observed outcome difference is reliable enough for decisionmaking.

Trial results should be discussed with the patient, and decisions regarding future treatment should take these results into account. There are numerous ways to present trial results to patients, and it is important to recognize that all patients will not benefit equally from the same presentation method. Providing the same information both graphically and numerically will allow the patient to explore the results in the manner that is most meaningful to him/her.[21] The researcher should always be responsive to the patient's needs, preferences, and goals, and promote shared decisionmaking. At the same time, the cumulative gain of knowledge from participating in n-of-1 trials can be emphasized. For example, even if a small effect is seen with one intervention, this may be a stepping stone to greater improvement. Scenarios for how to handle ambiguous results or results that do not favor treatment continuation should have been discussed during the trial planning phase. Please see Chapter 4 for more detailed information regarding the analysis and presentation of n-of-1 trial results.

Finally, all support and training must be provided in a way that minimizes the time commitment necessary for a patient to participate in an n-of-1 trial. The demands on a patient's time need to be realistic.[18]

Provider Users

Though researchers interested in conducting n-of-1 studies may interact directly with patients in conducting trials, researchers will often be partnering with both patients and providers in experimentation. Historically clinicians have played a central role in the execution of n-of-1 trials, either by carrying out the n-of-1 trial directly or through working with patients to interpret and implement trial results in cases where the trial was conducted by an n-of-1 consult service (either locally or remotely).[9] Therefore, it is important that researchers address the needs of the clinician when designing systems to support n-of-1 trials.

In order to successfully execute n-of-1 trials at any scale, researchers must convince providers that the benefits outweigh the inconvenience. In addition, researchers must provide adequate training and support to integrate n-of-1 trials into the clinical workflow at the lowest possible transactional costs for providers.

Engagement and Motivation

When attempting to engage clinicians in conducting n-of-1 trials, researchers should target clinics and specialties that will find them the most useful. Traditionally, primary care providers in fields such as family medicine, general internal medicine, and pediatrics as well as specialty clinicians who manage patients with chronic diseases have participated in n-of-1 trials.[6,10] However, researchers are likely to be successful engaging providers from a broader range of fields as long as those clinicians feel that they have a management problem that needs to be solved and that n-of-1 methods will help in that solution (see Chapter 1).

The key to motivating clinicians to participate in n-of-1 trials is to persuade them of the benefits. Researchers are encouraged to highlight the following advantages of n-of-1 trials:

1. **Enables truly personalized evidence-based medicine (EBM):** According to the Institute of Medicine, the practice of EBM means that to the greatest extent possible, health care decisions are grounded on a reliable evidence base, account for individual variation in patient needs, and support creation of new knowledge regarding clinical effectiveness.[22] N-of-1 trials have the potential to provide the highest strength of evidence for making individualized treatment decisions, in that they provide truly personalized evidence of clinical effectiveness.[10,23] While clinical guidelines and standardized care algorithms have emerged as methods to facilitate consistent application of evidence in practice and to eliminate variation among providers, these shared baselines are still meant to allow for variation to accommodate differences in patient needs and preferences.[24] N-of-1 trials provide a means for supporting such expected individual variation with sound evidence.
2. **Improves relationships between patients and providers:** Engaging in an n-of-1 trial facilitates collaboration between patients and providers. N-of-1 trials allow patients to participate more comprehensively in their own care—promoting self-management, greater insight into the disease, and personal

engagement in their own health.[4,13] N-of-1 trials increase communication between the patient and the clinician and support shared decisionmaking above and beyond what traditionally exists in current care models.[10] The process of custom designing n-of-1 trials—selecting interventions, determining which outcomes to measure, specifying stopping rules, and agreeing on the desired effect sizes—establishes a genuine two-way partnership between patients and providers.[10] In addition, n-of-1 trials can be beneficial when there is disagreement between patient and provider regarding the best approach to treatment.[6,25]

3. **Provides more precise answers about how to select among treatment options:** N-of-1 trials increase the precision of clinical decisionmaking in a number of ways. Current methods of trial-and-error prescribing among clinicians and self-experimentation among patients have little rigor and may provide misleading results. In most instances, insufficient data are collected to provide clear evidence of effectiveness, and even when data are collected, the apparent effectiveness of a treatment in the short term may only be the result of random variation in the patient's symptoms or the effects of uncontrolled external factors. Additionally, treatments that initially produce subtle improvements may be abandoned before their efficacy is ever appreciated. The more comprehensive, concrete, personalized information that surfaces from n-of-1 trials (e.g., from daily symptom diaries) provides a better understanding of symptom patterns and frequency that allows for deeper insight into the condition and overall better management.[13] By making clinical uncertainty explicit and using a rigorous design that includes randomization or counterbalancing, multiple crossover treatment periods, systematic outcome assessment, and blinding (if possible), n-of-1 trials enable providers to make informed decisions about the effects of various treatments in a way that reduces cognitive bias, one of the main threats of informal experimentation.[10] Studies of n-of-1 trials in medicine show increased provider confidence in their treatment decisions. In one of the original series of n-of-1 trials published by Guyatt et al., 84 percent of completed trials provided a definitive clinical answer, and physicians reported a high level of confidence in their treatment decisions in over 80 percent of trials.[7]

A range of approaches can be used to market these benefits of n-of-1 trials to providers. Strategies that have previously been used to promote n-of-1 trials among care providers include newsletters, professional media, Web sites, and presentations at clinical meetings.[8] Engaging a local physician champion within the clinic or unit may also be a helpful strategy for engaging other clinicians. However, researchers are encouraged to think broadly and creatively when identifying forums and methods to engage with care providers around n-of-1 trials.

Training and Support

Although most clinicians have had formal exposure to the concepts of EBM as well as research and trial design, many providers may not be familiar with n-of-1 methods. Researchers who hope to engage providers in n-of-1 trials need to educate clinicians about their basic features, emphasizing the validity and safety of the approach as well as specific issues around analysis such as display of data and how patients and clinicians determine whether a particular intervention has resulted in improvement (see Chapter 4).[10] Providers may also be particularly interested in more generalizable results, which can be achieved by aggregating data across n-of-1 trials (see Chapter 4 for additional information about aggregating n-of-1 trials). Researchers who wish to create large-scale n-of-1 trial systems should consider developing a scalable n-of-1 curriculum (e.g., online tutorials) to educate providers and other potential end-users about the basics of n-of-1 methods.

In addition to providing adequate training, researchers must also offer tools to support the shared decisionmaking that is central to n-of-1 trials. It is also important for researchers to support the major paradigm shift that accompanies n-of-1 trials in the current care delivery systems, including intensive patient-provider collaboration, explicit recognition of clinical uncertainty, and a more formal structure for experimentation beyond current models of informal experimentation.[10] In addition, researchers should help clinicians engage more intensively with patients, particularly with the growing number of self-experimenters (e.g., the Quantified Self community) who have developed their own expertise in data tracking and hypothesis testing.[4] Development of tools to support shared decisionmaking in n-of-1 trials is an area in need of additional research.

Researchers must also provide support to providers to allow them to conduct n-of-1 trials effectively within the setting of their typical clinical workflow and with minimal demands on time; setting realistic expectations regarding time and resources is also important. This support should be flexible enough to cover a range of engagement needs, from those clinicians who would like a great deal of autonomy and flexibility in designing and conducting trials to those who want a more prescriptive approach. Less intensive support could be provided by researchers in the form of a tutorial service that serves mainly educational needs.[7] More intensive support to guide clinicians through study design and execution, including identifying a study question, selecting outcome measures, designing the trial, and analyzing the data, could be provided in numerous ways; however, to date, this type of support has typically been provided by n-of-1 consult services.[6,9] For example, the n-of-1 consult service described by Larson et al. consisted of a core research group including a general internist, clinical pharmacist, family practitioner, and biostatistician. This service provided the key support functions for n-of-1 trials, including assistance with randomization and blinding as necessary, as well as with Institutional Review Board application preparation (see Chapter 2), general study design, and analysis. Having been offered this type of support, 85 percent of physicians indicated they experienced little or no inconvenience in referring patients to the n-of-1 service, and 77 percent reported that they spent no extra time or effort on their patients' participation in the trials.[6] Nikles et al. have also published their experience with a successful n-of-1 consult service delivered remotely across Australia.[8] This service consisted of centralized administrative support facilitated by use of mail, telephone, and electronic communication as well as standardized kits containing all the necessary supplies and information for an n-of-1 trial, including randomized doses, symptom diaries, and instructions. Kits were mailed directly to treating physicians, who were provided with trial results at the completion of the trial.[8,15] Nikles et al. found that >80 percent of physicians reported that they would order more n-of-1 trial kits and believed that n-of-1 trials were useful and worth the time commitment.[8] There has also been one model of a commercial n-of-1 trial service called Opt-e-Script. Although this venture was unsuccessful, it used the same combination of prefabricated kits (including blinded treatments and questionnaires) and analytic support.[1] The key to these successful n-of-1 consult services has been dynamic leadership, a multidisciplinary team, and a focused investment of resources.[1] While support for conducting n-of-1 trials has typically been offered through consult services, researchers are encouraged to identify other (perhaps more sustainable) ways of providing similar support to providers who are interested in executing n-of-1 trials.

The support offered by researchers to enable providers to conduct n-of-1 trials must address what has persistently been one of the most prominent barriers—increased time demands associated with n-of-1 trials. When discussing barriers to n-of-1 trials, physicians have mainly reported logistical concerns, particularly the administrative time demands in addition to time already spent on patient care.[10] Time-tracking data from some of the original n-of-1 consult services revealed that an average of 16.75 hours was spent on any one individual trial (total of the time spent by the entire team), half of which was spent on trial preparation.[6] Though this is not an estimate of the time required from an individual provider, and it is likely an overestimate for today's researcher, it underscores the need to develop systems that take advantage of new technologies to conduct n-of-1 trials in a way that minimizes the time required to participate (see Chapter 5). In addition, n-of-1 trials represent a different way of engaging and collaborating with patients—an approach that is not currently easily supported within the setting of the traditional clinical encounter. In fact, time-tracking data reflect that only one-third of the time spent on n-of-1 trials involved patient visits.[6] Researchers need to develop methods to support other types of patient-provider encounters that may result from engagement in design and execution of an n-of-1 trial, including out-of-office encounters and email correspondence. Creating methods that allow for compensation of provider time spent conducting n-of-1 trials is another important aspect of reducing transactional barriers that could potentially be achieved through Medicare or traditional insurance reimbursement (see Chapter 3 for additional details on financing of n-of-1 trials).

Development of new strategies and tools to educate and support clinicians in the design, execution, and analysis in n-of-1 trials that are fast, flexible, and inexpensive is one of the key areas of research necessary to move n-of-1 trials further into mainstream clinical practice.

Collaboration Among Users

Researchers must also design systems to allow for active collaboration among all types of users, including providers, patients, and researchers. This type of active collaboration across user groups is necessary to optimally facilitate engagement and provide adequate support for users. Collaboration among patients (e.g., through online communities) can allow them to learn from the experiences of other patients and provide opportunities to improve both the tools used to conduct n-of-1 trials (outcome measures, tracking tools, study protocols, etc.) as well as their own decisionmaking by incorporating results of prior trials either informally through review of available records, or formally through Bayesian analyses. Collaboration between various provider types such as physicians, pharmacists, dietitians, and psychologists is necessary to fully execute all aspects of an n-of-1 trial. For example, pharmacists can assist with packaging blinded medications, preparing randomization schedules, and providing information on the time of onset to action and washout periods of a particular drug in order to inform study design.[9] Psychologists can help researchers with aspects from patient adherence to data reporting and also with interventions targeting behavioral modifications.

Collaboration among user groups can bridge the gap between independent efforts by patients to engage in self-experimentation and efforts by researchers and health care providers to generate and apply evidence within the traditional health care delivery system. Examples of leveraging new types of technology-enabled collaborations across user groups to advance research, and useful models for researchers interested in advancing n-of-1 methods, include DIYGenomics, which is a not-for-profit research organization that focuses on crowd-sourced health studies and mHealth development, along with Genomera, which handles online study operations and engagement with patient communities.[26,27] These types of multidisciplinary online communities could ultimately become self-sustaining repositories for n-of-1 protocols and results in a way that advances the field of n-of-1 methods in health care.

Conclusion

N-of-1 trials have the potential to provide the highest level of evidence-based medicine for the individual, but are currently underutilized due to barriers at both the patient and provider levels. In this chapter, we offer methods to address these barriers with patients and providers in order to better engage, train, and support both parties. Reducing transactional costs for all participants may help increase the use of n-of-1 trials in clinical practice. It is recommended that researchers utilize the checklist below when addressing these barriers with patients and providers.

Checklist

Guidance	Key Considerations	Check
Engage patients through emphasizing the purpose and potential of n-of-1 trials	• Stress the potential to offer personal gain to participants, including greater insight into disease, improved self-management, and improvement in symptoms and quality of life not otherwise achieved with current plan. • Highlight flexibility and improved collaboration with providers. • Use a variety of social marketing approaches to target patients.	❐
Provide patients with basic education about n-of-1 methods	• Emphasize shared decisionmaking in the study design process. • Use patient-friendly outcome assessments and recording tools.	❐
Engage providers by emphasizing the purpose and potential of n-of-1 trials	• Emphasize ability to practice true personalized EBM, improved relationships with patients, and precision in decisionmaking. • Use a variety of marketing approaches to target physicians. • Address concerns about burden of trials and set realistic expectations.	❐
Provide clinicians with a basic education about n-of-1 methods	• Emphasize design concepts, validity, and safety, • Consider developing a scalable online n-of-1 curriculum to provide broader education.	❐
Provide user-friendly support tools to facilitate n-of-1 study execution and decrease time demands	• Tools should be directed specifically at improving shared decisionmaking in designing, executing, and interpreting results from n-of-1 trials. • Tools to support clinicians in the design, execution, and analysis in n-of-1 trials should always be directed toward making the process expeditious, flexible, practical, and economical.	❐
Design systems that encourage collaboration among all user types	• Explore the potential of online communities.	❐

References

1. Kravitz RL, Duan N, Niedzinski EJ, et al. What ever happened to N-of-1 trials? Insiders' perspectives and a look to the future. The Milbank Quarterly. Dec 2008;86(4):533-555.
2. Nikles CJ, Clavarino AM, Del Mar CB. Using n-of-1 trials as a clinical tool to improve prescribing. The British Journal of General Practice: The Journal of the Royal College of General Practitioners. Mar 2005;55(512):175-180.
3. Quantified Self. Self Knowledge Through Numbers. http://quantifiedself.com. Accessed November 4, 2011.
4. Mehta R. The Self-Quantification Movement–Implications for Healthcare Professionals. SelfCare. 2011;2(3):87-92.
5. Sackett DL. Clinician-trialist rounds: 4. why not do an N-of-1 RCT? Clinical trials. Jun 2011;8(3):350-352.
6. Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a 2-year period. Journal of the American Medical Association. Dec 8 1993;270(22):2708-2712.
7. Guyatt GH, Heyting A, Jaeschke R, et al. N of 1 randomized trials for investigating new drugs. Controlled Clinical Trials. Apr 1990;11(2):88-100.
8. Nikles CJ, Mitchell GK, Del Mar CB, et al. An n-of-1 trial service in clinical practice: testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. Pediatrics. Jun 2006;117(6):2040-2046.
9. Guyatt G, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. Canadian Medical Association Journal (journal de l'Association medicale canadienne). Sep 15 1988;139(6):497-503.
10. Kravitz RL, Paterniti DA, Hay MC, et al. Marketing therapeutic precision: Potential facilitators and barriers to adoption of n-of-1 trials. Contemporary Clinical Trials. Sep 2009;30(5):436-445.
11. Kent MA, Camfield CS, Camfield PR. Double-blind methylphenidate trials: practical, useful, and highly endorsed by families. Archives of Pediatrics and Adolescent Medicine. Dec 1999;153(12):1292-1296.
12. Scuffham PA, Yelland MJ, Nikles J, et al. Are N-of-1 trials an economically viable option to improve access to selected high cost medications? The Australian experience. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research. Jan-Feb 2008;11(1):97-109.
13. Brookes ST, Biddle L, Paterson C, et al. "Me's me and you's you": Exploring patients' perspectives of single patient (n-of-1) trials in the UK. Trials. 2007;8:10.
14. Gabler NB, Duan N, Vohra S, et al. N-of-1 trials in the medical literature: a systematic review. Medical Care. Aug 2011;49(8):761-768.
15. Nikles CJ, Yelland M, Glasziou PP, et al. Do individualized medication effectiveness tests (n-of-1 trials) change clinical decisions about which drugs to use for osteoarthritis and chronic pain? American Journal of Therapeutics. Jan-Feb 2005;12(1):92-97.
16. Yelland MJ, Nikles CJ, McNairn N, et al. Celecoxib compared with sustained-release paracetamol for osteoarthritis: a series of n-of-1 trials. Rheumatology. Jan 2007;46(1):135-140.
17. Woodfield R, Goodyear-Smith F, Arroll B. N-of-1 trials of quinine efficacy in skeletal muscle cramps of the leg. The British Journal of General Practice: The Journal of the Royal College of General Practitioners. Mar 2005;55(512):181-185.
18. Lillie EO, Patay B, Diamant J, et al. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? Personalized Medicine. Mar 2011;8(2):161-173.
19. Jaeschke R, Adachi J, Guyatt G, et al. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. Journal of Rheumatology. Mar 1991;18(3):447-451.
20. Jaeschke R, Singer J, Guyatt GH. A comparison of seven-point and visual analogue scales. Data from a randomized trial. Controlled clinical trials. Feb 1990;11(1):43-51.
21. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. Journal of the National Cancer Institute. Oct 5 2011;103(19):1436-1443.
22. The Learning Healthcare System: Workshop Summary. Paper presented at: Institute of Medicine. 2007; Washington, D.C.

23. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. Journal of the American Medical Association. Sep 13 2000;284(10):1290-1296.

24. James B. Quality Improvement Opportunities in Health Care–Making It Easy to Do It Right. JMCP. 2002;8(5):394-399.

25. Kamien M. The use of an N-of-1 randomised clinical trial in resolving therapeutic doubt. The case of a patient with an 'attention disorder.' Australian Family Physician. Jul 1998;27 Suppl 2:S103-105.

26. www.diygenomics.org/about.php.

27. http://genomera.com.

Authors

Chapter 1. Introduction to N-of-1 Trials: Indications and Barriers

Richard L. Kravitz, M.D., M.S.P.H.
Department of Internal Medicine
University of California, Davis
Davis, CA

Naihua Duan, Ph.D.
Department of Psychiatry
Columbia University
New York, NY

Sunita Vohra, M.D.
Department of Pediatrics
University of Alberta
Edmonton, Alberta, Canada

Jiang Li, M.D.
Evidence-Based Medicine Center
School of Basic Medical Sciences
Lanzhou University
Lanzhou, Gansu, China

The DEcIDE Methods Center N-of-1 Guidance Panel
(please see list of panel members below)

Chapter 2. An Ethical Framework for N-of-1 Trials: Clinical Care, Quality Improvement, or Human Subjects Research?

Salima Punja, B.Sc., Ph.D. Candidate
CARE Program (Department of Pediatrics) and Department of Medicine
Faculty of Medicine and Dentistry
University of Alberta
Edmonton, Alberta, Canada

Ian Eslick, Ph.D.
MIT Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA

Naihua Duan, Ph.D.
Professor Emeritus
Division of Biostatistics
Department of Psychiatry
Columbia University
New York, NY

Sunita Vohra, M.D., M.R.C.P.C., M.Sc.
Director, CARE Program
Professor, Department of Pediatrics
Faculty of Medicine and Dentistry
University of Alberta
Edmonton, Alberta, Canada

The DEcIDE Methods Center N-of-1 Guidance Panel
(please see list of panel members below)

Chapter 3. Financing and Economics of Conducting N-of-1 Trials

Wilson D. Pace, M.D.
Green-Edelman Chair for Practice-based Research and Professor of Family Medicine
University of Colorado Denver School of Medicine
Director, American Academy of Family Physicians National Research Network

Eric B. Larson, M.D., M.P.H.
Vice President for Research, Group Health Cooperative
Executive Director, Group Health Research Institute
Professor of Medicine and Health Services, University of Washington
Seattle, WA

Elizabeth W. Staton, M.S.T.C.
Instructor of Family Medicine, University of Colorado School of Medicine
Writer, American Academy of Family Physicians National Research Network

The DEcIDE Methods Center N-of-1 Guidance Panel
(please see list of panel members below)

Chapter 4. Statistical Design and Analytic Considerations for N-of-1 Trials

Christopher H. Schmid, Ph.D.
Professor of Biostatistics
Department of Biostatistics
Center for Evidence Based Medicine
School of Public Health
Brown University
Providence, RI

Naihua Duan, Ph.D.
Professor Emeritus
Division of Biostatistics
Department of Psychiatry
Columbia University
New York, NY

The DEcIDE Methods Center N-of-1 Guidance Panel
(please see list of panel members below)

Chapter 5. Information Technology (IT) Infrastructure for N-of-1 Trials

Ian Eslick, Ph.D.
MIT Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA

Ida Sim, M.D., Ph.D.
Co-Director, Biomedical Informatics
Clinical and Translational Sciences Institute
Professor of Medicine
University of California San Francisco
San Francisco, CA

The DEcIDE Methods Center N-of-1 Guidance Panel
(please see list of panel members below)

Chapter 6. User Engagement, Training, and Support for Conducting N-of-1 Trials

Heather C. Kaplan, M.D., M.S.C.E.
Assistant Professor
Perinatal Institute
James M. Anderson Center for Health Systems Excellence
Cincinnati Children's Hospital Medical Center
Cincinnati, OH

Nicole B. Gabler, Ph.D., M.P.H., M.H.A.
Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine
University of Pennsylvania
Philadelphia, PA

The DEcIDE Methods Center N-of-1 Guidance Panel
(please see list of panel members below)

The DEcIDE Methods Center N-of-1 Guidance Panel

Naihua Duan
Division of Biostatistics
Department of Psychiatry
Columbia University

Ian Eslick
MIT Media Laboratory

Nicole B. Gabler
Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine
University of Pennsylvania

Heather C. Kaplan
Perinatal Institute and James M. Anderson Center for Health Systems Excellence,
Cincinnati Children's Hospital Medical Center
Department of Pediatrics, University of Cincinnati College of Medicine

Richard L. Kravitz
Department of Internal Medicine
University of California, Davis

Eric B. Larson
Group Health Research Institute

Wilson D. Pace
Department of Family Medicine
University of Colorado Denver

Christopher H. Schmid
Center for Evidence-Based Medicine and Department of Biostatistics
Brown University

Ida Sim
Division of General Medicine
University of California San Francisco

Sunita Vohra
Department of Pediatrics
University of Alberta

Suggested Citations

Please cite this User's Guide as follows:

Kravitz RL, Duan N, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014.

Please cite individual chapters as follows:

Chapter 1:

Kravitz RL, Duan N, Vohra S, Li J, the DEcIDE Methods Center N-of-1 Guidance Panel. Introduction to N-of-1 Trials: Indications and Barriers. In: Kravitz RL, Duan N, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014: Chapter 1, pp. 1-11.

Chapter 2:

Punja S, Eslick I, Duan N, Vohra S, the DEcIDE Methods Center N-of-1 Guidance Panel. An Ethical Framework for N-of-1 Trials: Clinical Care, Quality Improvement, or Human Subjects Research? In: Kravitz RL, Duan N, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014: Chapter 2, pp. 13-22.

Chapter 3:

Pace WD, Larson EB, Staton EW, the DEcIDE Methods Center N-of-1 Guidance Panel. Financing and Economics of Conducting N-of-1 Trials. In: Kravitz RL, Duan N, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014: Chapter 3, pp. 23-32.

Chapter 4:

Schmid CH, Duan N, the DEcIDE Methods Center N-of-1 Guidance Panel. Statistical Design and Analytic Considerations for N-of-1 Trials. In: Kravitz RL, Duan N, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014: Chapter 4, pp. 33-53.

Chapter 5:

Eslick I, Sim I, the DEcIDE Methods Center N-of-1 Guidance Panel. Information Technology (IT) Infrastructure for N-of-1 Trials. In: Kravitz RL, Duan N, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014: Chapter 5, pp. 55-70.

Chapter 6:

Kaplan HC, Gabler NB, the DEcIDE Methods Center N-of-1 Guidance Panel. User Engagement, Training, and Support for Conducting N-of-1 Trials. In: Kravitz RL, Duan N, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014: Chapter 6, pp. 71-81.

The following Web site should be added to each citation:

www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm.